# On Removing the Condorcet Influence from Pairwise Elections Data

Abhijit Chandra, Sunanda Roy

Working Paper No. 10033
September 2010

IOWA STATE UNIVERSITY
Department of Economics
Ames, Iowa, 50011-1070

# On removing the Condorcet influence from pairwise elections data

Abhijit Chandra
Department of Mechanical Engineering, Iowa State University
email: achandra@iastate.edu, Tel: (515) 294 4834
and
Sunanda Roy *
Department of Economics, Iowa State University
email: sunanda@iastate.edu, Tel: (515) 294 4197

July 2, 2010

**Abstract**

Recent developments in voting theory show that Condorcet profiles embedded in electorates are responsible for conflicts between pairwise voting methods and for reversals of rankings under positional methods whenever candidates are dropped or added. Because of the strong symmetry of the rankings of the candidates within these profiles, it can be argued that Condorcet profiles represent complete ties between the candidates so far as election outcomes are concerned. Hence removing their influence from pairwise tallies should not matter

*The authors gratefully acknowledge comments from Donald G. Saari, George A. Hazelrigg and an anonymous referee on an earlier version of the work. All remaining errors are ours. Some of the results presented here are extensions of work supported by the U.S. National Science Foundation under Grant No. CMMI-0640826 and CMMI-0900093. Chandra gratefully acknowledges this support. Any opinions, conclusions or recommendations expressed are those of the authors and do not necessarily reflect views of the sponsoring agencies.

and moreover is justified because of the distortions they induce. The paper discusses a method of removing or reducing the influence of Condorcet profiles from pairwise elections data.

JEL Numbers: D70, D71, D72

Keywords: Condorcet cycles, voting paradoxes, pairwise election methods, Borda Count

# 1 Introduction

Two aggregation problems or paradoxes have received much attention in social choice theory. The first is that different election methods based on pairwise comparisions of the candidates[1] often lead to widely different election outcomes, raising concerns about whether election outcomes reflect the rather arbitrary choice of a procedure or the collective "will of the people". The second problem is the frequent reversal of rankings that happens under a given positional method, when candidates are dropped from or added to a set, although voters' preferences remain fixed. The second problem opens the door to election manipulation by interested parties.

Recent developments in voting theory have traced the cause of these problems to the presence of *Condorcet profiles* embedded within a given electorate[2]. The set of rankings within a Condorcet profile demonstrate a strong symmetry across the candidates - each candidate is placed in each position by exactly the same number of voters. Thus, for this bloc of the electorate, the candidates are completely tied. However, when pairwise scores are formed from these rankings across the candidates, cycles of intransitive pairwise rankings emerge. If the entire electorate is one Condorcet profile, an aggregate ranking using the pairwise scores is not possible - a situation known as Condorcet's paradox. As is more usual however, if such a profile is a small portion of the electorate, problems surface in the form of disagreements between different pairwise voting methods and of ranking reversals when candidates are dropped or added, under a given positional method.

As an illustrative example, consider an election contested by three candidates $A$, $B$ and $C$. There are 126 voters, 19 of whom have the preference $A > B > C$, 26 have $A > C > B$, 6 have $C > A > B$, 33 have $C > B > A$, 26 have $B > C > A$ and the remaining 16 have $B > A > C$. Using standard convention, the voters profile is given by $(19, 26, 6, 33, 26, 16)$. The pairwise tallies are, $(A : B) = (51 : 75)$, $(A : C) = (61 : 65)$, and $(B : C) = (61 : 65)$. Thus $C$ is the Condorcet winner. The Borda Count for the candidates are $A = 112$, $B = 136$, and $C = 130$ and $B$ is the Borda winner. The conflict between the Condorcet and the Borda outcomes can be attributed to the presence of a bloc of 24 voters, of whom 8 favor the ranking $A > C > B$, another 8 favor $C > B > A$ and the

[1]Popular legislative procedures such as amendment and successive procedures are based on pairwise rankings of candidates. See Rasch (1995), for a survey. Opinion polls are frequently conducted as match ups between pairs of the main contenders. Pairwise methods include the 18th century Condorcet approach(es) and more recent innovations such as Copeland's method (1951), Black's method (1958), Kemeny's rule (1959) and Slater's method (1961). Further, the positional method of Borda Count can be constructed from pairwise election tallies across the candidates.

[2]The seminal papers are Saari (1999), (2000a) and (2000b).

remaining 8 voters favor $B > A > C$ - that is the profile component $(0,8,0,8,0,8)$ embedded in the larger profile $(19,26,6,33,26,16)$. Further, the profile $(0,8,0,8,0,8)$ is responsible for reversals in ranking under any fixed positional method, when candidates are dropped or added. For instance, under plurality the aggregate ranking is $A > B > C$. If however $A$ is dropped, on the remaining two candidates $B$ and $C$ the aggregate ranking becomes $C > B$. Similarly if $C$ is dropped, the aggregate ranking becomes $B > A$.

In this paper, we attempt a partial solution to these aggregation problems, in tune with a general recognition in the literature that voting methods have a different and somewhat more limited goal, compared to social choice functions.[3] The goal of a voting method is the practical and more limited one of consistently designating a winner from multiple candidates. The goal of a social choice function is the broader and more complicated one of consistently designating a socially optimal ranking across the candidates. Given the perfect symmetry of rankings within a Condorcet profile, one can argue that within this bloc of voters, the candidates are completely tied. Therefore such profiles should be excluded and the election outcome be determined by the rest of the electorate. Note, we do not claim that Condorcet profiles be excluded to determine society's *most preferred ranking* of the $n$ candidates. Condorcet profiles can be interpreted as ties across candidates but not across rankings.[4] As elections are essentially devices to break ties across candidates, excluding such profiles to determine an election outcome is justified.[5]

Assuming that Condorcet profiles may be justifiably excluded to determine an election outcome, we present here a method to isolate and remove or reduce the component contributed by Condorcet profiles to pairwise elections data, so that the election outcome is determined by the rest of the electorate. Recent results from geometric voting theory provide the foundation for our method.

Paradoxes due to the presence of Condorcet profiles arise because pairwise methods by their

[3]The most recent such views are articulated in Balinski and Laraki, 2007, 2009. Also see Moulin(2003).

[4]Balinski and Laraki (2009) explains why, with an example. Suppose in an electorate of 1000 voters and three candidates, 333 voters prefer $A > B > C$, 333 voters prefer $B > C > A$, 333 voters prefer $C > A > B$ and 1 voter prefers $A > C > B$. According to the first 999 voters, the candidates are all tied and hence in an election the last voter can break the tie to elect the winner $A$. However, $A > C > B$ is clearly not society's optimal ranking of the three candidates as only one voter likes it.

[5]Alternatively, an unified approach to voting and social choice functions may look for aggregation methods that simultaneously satisy a *weaker* set of ideal requirements or a fewer set of requirements - specifically, we may not require that such aggregation methods simultaneously satisfy the *independence of irrelevant alternatives* (IIA), *universality* and *unanimity* properties. For a survey, see Hodge and Klima (2005).

nature do not use (nor indeed do positional methods over subsets of candidates) the complete information provided by the multilateral and transitive rankings within the profile. Instead they use partial information in the form of selective binary components of these rankings. Under our proposed method, the given pairwise scores are revised in a way that restores some of this lost information. An initial (given) pairwise score difference between two candidates $i$ and $j$ is adjusted using all possible pairwise scores for $i$ and $j$ against other candidates, in an iterative fashion. We show that the iterative procedure reduces the weight contributed by the Condorcet profiles in each pairwise score by an *uniform* amount and increases the weight contributed by other types of profiles, also by an *uniform* amount.[6] Thus embedded profiles other than the Condorcet profiles, determine the election outcome when any pairwise method or the Borda count is used on the *revised* pairwise scores.

A main result (Theorem 1) shows that if pairwise scores for all candidates *across whom voters' preferences are defined* are available, our method can completely eliminate the Condorcet components from the individual pairwise scores. We describe this scenario as a *complete field* election. Under this (ideal) scenario, all pairwise election methods applied on the revised scores will agree and will further agree for any subset of candidates (Corollary 1). In particular, the Condorcet approach(es) and the Borda Count will agree on the revised pairwise scores for any subset of candidates.

A less than ideal scenario is an *incomplete field* election, under which pairwise scores for a *smaller* subset of candidates are available compared to the set over which voters' preferences are defined.[7] Although under this scenario, our method does not completely eliminate all Condorcet components, the second main result (Theorem 3) of the paper shows that a Borda Count on the revised scores for each candidate has a smaller Condorcet component compared to the Borda Count obtained from the initial unrevised scores. Further, we also provide conditions (Theorem 4) under which the Condorcet components from the individual pairwise scores for all pairs can be reduced compared to the initial levels.

Section 2 lays down the geometric voting theory framework used in the paper. Section 3

[6] As is well known and as we discuss in the next section, there is only one other type of profile which contributes to a pairwise score, namely, *Basic profiles*. Although the weight contributed by Basic profiles increase uniformly under the iterations, election outcomes are not distorted.

[7] Opinion poll match-ups for example, are conducetd across only a subset of the main contenders. Section 2 has more discussion.

presents our method. Section 4 presents the main results and an illustrative example from the US 2008 presidential primaries.

## 2 The geometric voting theory framework

Assume that a given set of voters have individually strict and transitive preferences over $n$ candidates named $A, B, \ldots N$ and indexed $i = 1 \ldots n$. There are $n!$ different rankings of these candidates. An electorate profile is a vector $p = (p_1 \ldots p_{n!}) \in \mathbf{R}_{+}^{n!}$ where $p_j$ represents the number of people in the given electorate with preferences given by the $j$th ranking of the candidates. A profile differential $p' \in \mathbf{R}^{n!}$ is the difference between two different profiles for an electorate of a given size.[8] A normalized profile differential is a vector $p \in \mathbf{R}^{n!}$ such that $\sum_{j=1}^{n!} p_j = 1$.

A *Basic profile differential* favoring candidate $i$, denoted $B_i^n$, is a profile which has one voter for each ranking that has $i$ top ranked, (-1) voter for each ranking that has $i$ bottom ranked and 0 voter for each ranking that has $i$ ranked somewhere in the middle.

To define a Condorcet profile, denote by (1) the reference ranking $A > B > C \ldots > N$. A *Condorcet n-tuple*, $c_{(1)}^n$ and its reversal set $\rho(c_{(1)})^n$, generated by reference ranking (1) are the sets of $n$ rankings given by,

Table 1:

| $c_{(1)}^n$ | $\rho(c_{(1)})^n$ |
|---|---|
| $A > B > C \ldots > N$ | $N > M > L \ldots > A$ |
| $B > C > D \ldots > A$ | $M > L > K \ldots > N$ |
| $C > D > E \ldots > B$ | $L > K > J \ldots > M$ |
| $\ldots$ | $\ldots$ |
| $N > A > B \ldots > M$ | $A > N > M \ldots > B$ |

A Condorcet profile $C_{(1)}^n$ associated with the reference ranking $(1)$, is a profile that has one voter for each ranking in $c_{(1)}^n$ and (-1) voter for each ranking in $\rho(c_{(1)}^n)$ and zero voter for each remaining rankings in the list of all rankings.

[8] A profile differential is often a more convenient concept to work with than a profile. Note that for an appropriate choice of a constant $a$, the profiles $p'$ and $p = p' + aK^n$, where $K^n \in \mathbf{R}_{+}^{n!}$ is a profile which has one voter for each possible ranking, has the same election outcome. Hence $p$ and $p'$ are analytically equivalent objects even though $p'$ has ”negative” voters.

Denote by $a_{ij}$ the normalized difference in the scores of candidates $i$ and $j$ in a pairwise election, that is $a_{ij}$ equals $i$'s voter tallies minus $j$'s voter tallies, divided by the total number of voters. Thus $-1 \leq a_{ij} \leq 1$ with a value of 0 indicating a tie between the two candidates, and values of -1 and + 1 indicating unanimous loss or win by candidate $i$ against candidate $j$. Further, $a_{ij} = -a_{ji}$. A vector $a = \{a_{ij}\}_{i,j=1\ldots N, i<j}$ of normalized pairwise score differentials defines a point in the cube $RC(n) \subset \mathbf{R}^{n_{C_2}}$, defined by the $n_{C_2}$ intervals $[-1,1]$.

Pairwise score differences, $a_{ij}$, are determined only by the Basic and Condorcet profiles within an electorate. Other types of profiles contribute nothing towards these values. Thus, the $\{a_{ij}\}$ vectors which constitute the data for our method, are totally determined by the Basic and Condorcet profiles within an electorate.[9]

A Basic profile $B_i^n$ generates a vector of pairwise score differentials in $RC(n)$ that is a multiple of the vector $T_i^n \in RC(n)$, which has $a_{i,j} = 1$ for all $j$ and $a_{j,k} = 0$ for all $j,k \neq i$.[10] The collection $\{T_i^n\}_{i=1}^n$ generated by the $n$ Basic profiles for the $n$ candidates, spans a $(n-1)$ dimensional subspace $\mathbf{T} \subset RC(n)$ called the *Transitivity plane*.

A complete characterization of the directional vectors in $RC(n)$ generated by all possible Condorcet profiles is difficult because the number of distinct Condorcet profiles in an $n$-candidate election equals $\frac{(n-1)!}{2}$. However, given any reference ranking which generates a Condorcet profile, a directional vector can be algorithmically constructed.[11]

The set of $\frac{(n-1)!}{2}$ directional vectors of pairwise score differentials (generated by the $\frac{(n-1)!}{2}$ distinct Condorcet profiles) span a subspace of dimension $n_{C_2} - (n-1) = (n-1)_{C_2}$ in $RC(n)$, a higher dimensional subspace compared to the Transitivity plane. For convenience, we describe this subspace as the *Condorcet subspace*, $\mathbf{C} \subset RC(n)$.

A profile $p$ has the *additive transitivity* property if, for any subset of $k$ out of $n$ candidates and any permutation of the indices, $\sum_{j=1}^{k-1} a_{j,j+1} = a_{1k}$. Basic profiles and hence their weighted sums

[9] Saari (2000a).

[10] Under $B_i^n$, $i$ defeats all $j$'s *unanimously* in pairwise contests. The pairwise tallies of candidates $(i,j)$ are $(n-1)! : \frac{-(n-1)!}{(n-2)}$. Add $\frac{(n-1)!}{(n-2)}K^2$ to this pairwise profile to see that $i$ wins unanimously. The other pairwise contests not involving $i$ are all ties.

[11] The pairwise tallies of $(i,j)$ for a Condorcet profile are given by $(n-2s):(2s-n)$ if $i$ is ranked $s$ candidates above $j$ in the reference ranking, where ither $n-2s$ or $2s-n$ is negative unless $n=2s$. This implies that the normalized pairwise score differentials of any pair, $a_{ij}$, in a Condorcet profile is 1 if $n-2s>0$, implying $i$ wins unanimously over $j$: it is -1 if $n-2s<0$, implying $j$ wins unanimously over $i$; or it is 0 if $n-2s=0$, implying both are tied.

which lie on the Transitivity plane satisfy the additive transitivity property.[12]

The paper studies two scenarios under which our method may be used. The first is a *complete field* election. Under this scenario, pairwise elections are assumed to have been held for all candidates across whom *voters' preferences are defined* and hence pairwise score differences for all pairs of candidates across whom voters' preferences are defined, are available. This may or may not be a reality. For instance in opinion polls, match ups are often conducted over a subset of the candidates (see the example in the last subsection), regarded as ”major contenders” by the organization, although voters' preferences may be defined over a larger set of candidates. Or for instance, candidates may drop out while the race is on but continue to feature in voters' preferences. Under these situations, pairwise score differences for all candidates over which voter's preferences are defined, are not available. We describe this as an *incomplete field* election scenario, to underline the fact that there are ”missing” candidates for whom pairwise scores are not available. More details on these two scenarios are provided later.

The following result from Saari (2000a) provides the foundation and rationale for our present attempt.

**Lemma 1** ***(Saari, 2000a)*** *The subspaces* **T** *and* **C** *are orthogonal to each other. Further,* $a \in RC(n)$ *has a unique decomposition* $a = a^T + a^C$*, where* $a^T \in \mathbf{T}$ *and* $a^C \in \mathbf{C}$.

Profile decompositions are not easily implementable for elections with more than four candidates, as the number of distinct Condorcet profiles gets very large quickly. Decomposition of pairwise score differences (Lemma 1) is easier to implement than profile decomposition but difficult nevertheless for a large number of candidates as this involves solving a large system of equations which may not have a simple readily identifiable structure. Further, Lemma 1 cannot be implemented under an incomplete field election scenario. Our method of extracting the Condorcet components from the pairwise scores is easily implementable and further, useful under incomplete field election scenarios.

[12] Saari (2000a).

# 3 The method

Condorcet profiles cause distortions in pairwise score differences because pairwise elections focus on selective binary components of a multilateral ranking to the exclusion of the rest of the ranking. Consider the Condorcet profile generated by the reference ranking $A > B > C > D$ in a 4-candidate election. A pairwise election of the $(B,C)$ pair forces us to consider only the part of each ranking in this profile, that deals with $B$ and $C$. It ignores the fact for example, that some rankings within this profile has $A$ placed before $B$ and $D$ placed after $C$. Thus, pairwise methods do not use a critical piece of information about the profile - in the complete field of 4 candidates, each candidate is supported in each position by exactly the same number of voters.

To remove or reduce distortions caused by this loss of information, our method seeks to *reverse* the above process. Knowing that the pairwise score differences between $B$ and $C$ considers only one binary component of a multilateral ranking, we use all the other pairwise scores differences (which similarly only considers other binary components of the same set of rankings) to adjust the pairwise score differences between $B$ and $C$. Our objective is to thereby recapture some of the spirit of the original multilateral rankings.[13]

Let $a_{ij}^{(0)}$ denote the given initial pairwise score differences in an election between $i$ and $j$. Denoting revised pairwise score differences after one iteration of the method as $a_{ij}^{(1)}$, the initial scores are revised according to the algorithm

$$a_{ij}^{(1)} = a_{ij}^{(0)} + CF. \sum_{k \neq i,j} (a_{ik}^{(0)} + a_{kj}^{(0)}), \forall i,j \tag{1}$$

where $CF$, a positive constant described as the *Confidence factor*, is chosen as a measure of the confidence we place on the "indirect" evidence regarding $i$ and $j$. The indirect evidence is the set of pairwise scores differences of $i$ and $j$ against all the other opponents. The constant $CF$ is never chosen to exceed 1 (and is in fact restricted to $[0, \frac{1}{2}]$, later in the paper), implying that indirect evidence are never considered as valuable as direct evidence, the pairwise score difference between $i$ and $j$ itself.[14]

[13]Aggregation problems of the type addressed in this paper are not unique to social choice and voting theory but are common in many applied sciences, such as in many areas of Engineering. A version of the method presented in this section and described as the "Reciprocal Coordination" method was developed by Chandra, Karra and Dorothy (2008) in the discipline of Mechanical Engineering.

[14]We may describe this indirect evidence used here as "level 1" indirect evidence, underlining the fact that the path

Multiple iterations with our method may be necessary under an incomplete field election. At the $(m+1)$th iteration, the revised scores are given by $a_{ij}^{(m+1)}$, where

$$a_{ij}^{(m+1)} = a_{ij}^{(m)} + CF. \sum_{k \neq i,j} (a_{ik}^{(m)} + a_{kj}^{(m)}), \forall i, j \tag{2}$$

Finally, the election winner is determined by using any specific pairwise method or by doing a Borda Count on the revised pairwise scores, $a_{ij}^{(m)}$. We show in Scetion 4 that the revised scores have a zero Condorcet component, under a complete election scenario, and a lower Condorcet component, under an incomplete election scenario, compared to the initial scores, $a_{ij}^{(0)}$. Hence, by using the revised scores to determine the winner, any of the standard pairwise methods or the Borda Count will demonstrate greater consistency compared to what they would by using the initial unrevised scores.

## 3.1 Decomposition and closed forms for revised scores

By Lemma 1, $a_{ij}^{(0)}$ can be decomposed into Transitive and Condorcet components, $a_{ij}^{(0,T)}$ and $a_{ij}^{(0,C)}$. Propositions I and II below, provide closed form expressions for the revised pairwise scores $a_{ij}^{(m)}$, as a function of $a_{ij}^{(0,T)}$, $a_{ij}^{(0,C)}$ and the parameters $n, CF$, and $m$, under both complete and incomplete field election scenarios. These expressions help us understand how the initial weights of these two components are affected by the iterations.

**Proposition 1** *Under a complete field election scenario,* $a_{ij}^{(m)} = (1+(n-2)CF)^m a_{ij}^{(0,T)} + (1-2CF)^m a_{ij}^{(0,C)}$.

*Proof*: By Lemma 1,

$$a_{ij}^{(1)} = a_{ij}^{(0,T)} + a_{ij}^{(0,C)} + CF. \sum_{k \neq i,j} (a_{ik}^{(0,T)} + a_{kj}^{(0,T)} + a_{ik}^{(0,C)} + a_{kj}^{(0,C)}) \forall i, j$$

By the additive transitivity property of all scores which lie on the Transitivity plane,

$$a_{ik}^{(0,T)} + a_{kj}^{(0,T)} = a_{ij}^{(0,T)}$$

from $i$ to $j$ is crossed via a single intermediate node $k$. The method can be extended to include "level $s$" indirect evidence, that is the path from $i$ to $j$ can be crossed via $s$ intermediate nodes. However the revised scores are the same, irrespective of how many intermediate nodes are used.

By the property of Condorcet profiles, for full elections,

$$\sum_{k\neq i} a_{ik}^{(0,C)} = 0 \Rightarrow \sum_{k\neq i,j} a_{ik}^{(0,C)} = -(a_{ij}^{(0,C)})$$

Finally note that $a_{kj}^{(0,C)} = -a_{jk}^{(0,C)}$. Hence $\sum_{k\neq i,j} a_{kj}^{(0,C)} = -\sum_{k\neq i,j} a_{jk}^{(0,C)} = a_{ji}^{(0,C)} = -a_{ij}^{(0,C)}$
Using all of these $a_{ij}^{(1)} = (1+CF(n-2))a_{ij}^{(0,T)} + (1-2CF)a_{ij}^{(0,C)} \forall i,j$ after one iteration.
For $m$ iterations, Proposition follows by method of induction. $\Delta$.

Next consider an incomplete field election scenario. Assume that pairwise scores for only $n$ candidates are available but that voters' preferences are defined over $n+1$ candidates.Also without loss of generality assume that the $(n+1)$the candidate is the "missing" candidate.

Under an incomplete field election scenario, the sum of the pairwise Condorcet components for each candidate $i$, over her current $n-1$ opponents is not zero.

Conceptually as before, the vector of pairwise scores (some elements of which are unknown) can be decomposed into a Transitive and a Condorcet component, but in an $(n+1)_{C_2}$ dimensional space. Thus for each $i$, $\sum_{j\neq i, j=1\ldots n+1} a_{ij}^{(0,C)} = 0$ but $\sum_{j\neq i, j=1\ldots n} a_{ij}^{(0,C)} \neq 0$. In particular, it follows that $\sum_{k\neq i,j} a_{ik}^{(0,C)} = -(a_{ij}^{(0,C)}) - (a_{i(n+1)}^{(0,C)})$ and $\sum_{k\neq i,j} a_{kj}^{(0,C)} = -(a_{ij}^{(0,C)}) + (a_{j(n+1)}^{(0,C)})$, where the last term on the right hand sides are attributable to the missing candidate.

Define coefficients $t^{(1)} = (1-2CF) + nCF$ and $t^{(m)} = ((1-2CF)+nCF)t^{(m-1)}$; $p^{(1)} = (1-2CF)$ and $p^{(m)} = (1-2CF)p^{(m-1)}$; and $s^{(1)} = CF$ and $s^{(m)} = ((1-2CF)+nCF)s^{(m-1)} + CF\,p^{(m-1)}$ for the $m$th iteration.

**Proposition 2** *Under an incomplete field election scenario,*

$$a_{ij}^{(m)} = a_{ij}^{(0,T)} t^{(m)} + a_{ij}^{(0,C)} p^{(m)} + (a_{j(n+1)}^{(0,C)} - a_{i(n+1)}^{(0,C)}) s^{(m)} \qquad (3)$$

*where the coefficients have the following closed forms:*

$$\begin{aligned} t^{(m)} &= ((1-2CF)+nCF)^m \\ p^{(m)} &= (1-2CF)^m \\ s^{(m)} &= ((1-2CF)+nCF)^{m-1}.\frac{1}{n}(\frac{((1-2CF)+nCF)^{m-1}-(1-2CF)^{m-1}}{(1-2CF)+nCF)^{m-2}}, \;\; for \;\; m>1 \\ &= CF, \;\; for \;\; m=1 \end{aligned}$$

*Proof*: After one iteration,

$$a_{ij}^{(1)} = (1+CF(n-2))a_{ij}^{(0,T)} + (1-2CF)a_{ij}^{(0,C)} + CF(a_{j(n+1)}^{(0,C)} - a_{i(n+1)}^{(0,C)})\forall i,j$$

Eqn ( 3) is true for $m = 1,2$ and by induction, for any $m$. The closed forms for $t^{(m)}$ and $p^{(m)}$ can be similarly derived.

A rearrangement of the terms in the expression for $s^{(2)}$ and the method of induction shows that,

$$s^{(m)} = (1-2CF+nCF)^{m-1}.CF(\sum_{s=0}^{m-1} \frac{p}{(1+(n-2)CF)^s}$$

which on simplification and substitution of $p = (1-2CF)$ yields,

$$s^{(m)} = (1-2CF+nCF)^{m-1}.\frac{1}{n}(\frac{((1-2CF)+nCF)^{m-1} - (1-2CF)^{m-1}}{(1-2CF)+nCF)^{m-2}}. \qquad \Delta$$

.

Under both scenarios our method uniformly stretches the Transitivity components of the initial pairwise score differences by the factor $(1+CF(n-2))^m$. However, as this stretch is uniform across all candidates, no distortion in election outcomes result from using our method, so far as the Transitive component is concerned. Election tallies for $(1+(n-2)CF)^m a_{ij}^{(0,T)}$ and $a_{ij}^{(0,T)}$ may differ under different pairwise methods but the ranking of the candidates (the outcome), remains the same. Also note that under both scenarios, the choice for $CF$ must be restricted to the interval $[0,1/2]$ so that the method does not *distort* the initial Condorcet components, $a_{ij}^{(0,C)}$, through the iterations.

The main results presented below characterize choices of $CF$ and $m$ that reduce the weight of $a_{ij}^{(0,C)}$ to zero or to a number below one (the initial weight) under both scenarios.

# 4 Results

## 4.1 Complete field elections

Assume that pairwise elections have been held for all pairs of candidates across whom voters preferences are defined and therefore pairwise score differences are available for all pairs of candidates. Denote by $a^{(m)}$, the revised vector of pairwise score differentials after $m$ iterations.

**Theorem 1** *Under a full field election, for $CF = 1/2$, $a^{(1)}$ lies in the Transitivity plane, that is it does not have a Condorcet component.*

*Proof*: For $CF = 1/2$ and $m = 1$, $a_{ij}^{(1)} = (1 + CF(n-2))a_{ij}^{(0,T)}$ from Proposition I. Hence theorem follows. $\Delta$.

Under a complete field election, our method eliminates the Condorcet components with just one iteration (multiple iterations acheive nothing additionally). As mentioned before, stretching the Transitivity components of the initial pairwise scores by $(1+\frac{1}{2}(n-2))$ does not matter because $(1+CF(n-2))a_{ij}^{(0,T)}$ and $a_{ij}^{(0,T)}$ yield the same ranking of the candidates, under any pairwise method. Further, as $a^{0,T} = \frac{1}{(1+CF(n-2))}a^{(1)}$ and as calculating $a^{(1)}$ is computationally simpler than doing a direct decomposition for large $n$, Theorem 1 provides an easier way to extract $a^{(0,T)}$ from $a^{(0)}$, compared to Lemma 1.

An election method that is unaffected by Condorcet components when used over the complete set of candidates, is the Borda Count. In other words, Borda rankings of all $n$ candidates using either $a^{(0)}$ or $a^{(0,T)}$ are identical. However these rankings are not in general consistent over subsets of these $n$ candidates because of the presence of Condorcet profiles. An important advantage of using $a^{(1)}$ is that as the latter does not have a Condorcet component, *all* pairwise methods, including the Borda Count, will agree for *any subset of candidates* on $a^{(1)}$. We assert this usefulness of $a^{(1)}$ in part 2 of the following Corollary.

**Corollary 1** *For a full field election,*

1. *Each of the three vectors of pairwise score differences $a^0$, $a^{0,T}$ and $a^{(1)}$ yield the same ranking of the candidates under the Borda Count.*

2. *On $a^{(1)}$, the election outcomes for all subsets of candidates agree under any pairwise method and including under Borda Count.*

Suppose that in an election, *after* voters have cast their votes for each pair, a candidate $E$ says that even if she is declared the winner she is unable to accept the office. Part 2 of the Corollary says that if all the pairwise scores, including for candidate $E$, are first revised using our method, any pairwise method, including the Borda Count will be able to pick a consistent winner from the remaining candidates. Thus, using the revised scores, any of these methods will now pick a candidate who would have been placed second if $E$ had been placed first or who would have been placed first if $E$ had not been placed first. Using the initial unrevised scores on the other hand, none of the methods including the Borda Count, would pick a consistent winner in this sense, in general.

A slightly different version of the same example serves to highlight the next set of results.

## 4.2 Incomplete field election

Under an incomplete field election, pairwise elections have not been held for all pairs of candidates over whom voter's preferences are defined. In terms of the previous example, the candidate $E$ decides not to seek the office and drops out of the race but *before* voters have expressed their opinion on all pairs involving her and every other candidate. Here, as in the previous example, she has influenced the votes for all other pairs not involving her, by virtue of the fact that she appears in voters' preferences. But unlike in the previous example, no direct information about how she matches up with every other candidate exists.

In terms of the general model, voters' preferences include a missing, $(n+1)$th candidate on whom pairwise scores are not available. Proposition 2 shows that iteration with our method introduces a new term $(a^{(0,C)}_{j(n+1)} - a^{(0,C)}_{i(n+1)})s^{(m)}$ involving Condorcet components attributable to the missing candidate, into the revised scores. From expression ( 3), it is also clear that for no value of $CF$ (and $m$) do both $p^{(m)}$ and $s^{(m)}$ vanish simultaneously - that is some Condorcet component remains in the revised pairwise score differences. The question is, can our method do better compared to a "standard threshold" for appropriate choices of $CF$ and $m$. First, we identify such a threshold.

Note that applying the Borda method on the available initial unrevised scores, leaves us with a Condorcet component as well - a component attributable to the missing candidate. The Borda Count for candidate $i$ on the unrevised scores is given by $\sum_{j\neq i} a^0_{ij} = \sum_{j\neq i} a^{0,T}_{ij} - a^{0,C}_{i(n+1)}$. We denote the deviation $(-a^{0,C}_{i(n+1)})$ from the transitivity component as the *threshold Borda deviation*. The (unknown) magnitude of $(-a^{0,C}_{i(n+1)})$ provides a threshold for our method, as this is the best we can do with the unrevised scores and using a standard method which is known to be uninfluenced by

Condorcet components when there are no missing candidates.

Next we characterize values of $CF$ and $m$ for which we can do better than the threshold Borda deviation using the Borda method on the revised scores. This is done in two steps.

Denote by $\hat{a}_{ij}^{(m)} = \frac{1}{t^{(m)}} a_{ij}^{(m)}$, the revised normalized pairwise score difference for the pair $(i, j)$, using our method. First, Theorem 2 shows that for $CF = 1/2$ and $m = 1$, a Borda Count on $\hat{a}_{ij}^{(m)}$ and on $a_{ij}^0$ yield the same outcome - the resulting deviation from the transitivity component using $\hat{a}_{ij}^{(m)}$ is the same as the threshold Borda Deviation using the unrevised scores $a_{ij}^0$. We designate $CF = 1/2$ as the *threshold CF* - the value of $CF$ which yields the threshold Borda deviation using our method. Theorem 3 then shows that by choosing $CF < 1/2$ and $m > 1$, we can do better.

**Theorem 2** *If (1)$CF = 1/2$ and $m \geq 1$ or if (2) $CF \in (0, 1/2]$ and $m = 1$, the Borda Count for candidate i on scores $\hat{a}_{ij}^{(1)}$ is the same as the Borda Count on scores $a_{ij}^0$. In particular the Borda deviations under either scenario, for the revised scores are the same as the Threshold Borda Deviations.*

*Proof*: Note that from expression ( 3)

$$\frac{1}{t^{(m)}} a_{ij}^{(m)} = a_{ij}^{0,T} + a_{ij}^{0,C} \frac{p^{(m)}}{t^{(m)}} + (a_{j(n+1)}^{0,C} - a_{i(n+1)}^{0,C}) \frac{s^{(m)}}{t^{(m)}} \tag{4}$$

1. Substituting for $CF = 1/2$ and simplifying yields

$$\hat{a}_{ij}^{(1)} = a_{ij}^{0,T} + \frac{1}{n}((a_{j(n+1)}^{0,C} - a_{i(n+1)}^{0,C})$$

For each $i$, summing over $j \neq i$ candidates, noting that $a_{j(n+1)}^{0,C} = -a_{(n+1)j}^{0,C}$ and $\sum_j a_{(n+1)j}^{0,C} = 0$, and simplifying, we have

$$\sum_{j \neq i} \hat{a}_{ij}^{(1)} = \sum_{j \neq i} a_{ij}^{0,T} - a_{i(n+1)}^{0,C}.$$

2. When $m = 1$,

$$\hat{a}_{ij}^{(1)} = a_{ij}^{0,T} + (\frac{1 - 2CF}{1 + (n-2)CF}) a_{ij}^{0,C} + (\frac{CF}{1 + (n-2)CF})((a_{j(n+1)}^{0,C} - a_{i(n+1)}^{0,C})$$

Following the same steps as above and simplifying, we show

$$\sum_{j \neq i} \hat{a}_{ij}^{(1)} = \sum_{j \neq i} a_{ij}^{0,T} - a_{i(n+1)}^{0,C}. \quad \Delta$$

Theorem 2 shows that when pairwise scores are missing, one iteration for any $CF \in (0, 1/2]$, is not enough to get better results than the Borda Count on unrevised scores. Neither do multiple iterations with $CF = 1/2$ produce better results. Clearly multiple iterations with $CF < 1/2$ are needed.

**Theorem 3** *For $m > 1$ and $0 < CF < 1/2$, for candidate i, the Borda deviation on the revised scores is strictly less than the threshold Borda deviation $| - a^{0,C}_{i,(n+1)}|$.*

*Proof*: Note that for $m > 1$ and $0 < CF < 1/2$, $\frac{p^{(m)}}{t^{(m)}} = (\frac{1-2CF}{1+(n-2)CF})^m$ and $\frac{s^{(m)}}{t^{(m)}} = \frac{1}{n}(1 - (\frac{1-2CF}{1+(n-2)CF})^{m-1})$. Substituting into ( 4), summing $\hat{a}^{(m)}_{ij}$ over $j$ and simplifying as before, we have

$$\sum_{j \neq i} \hat{a}^{(m)}_{ij} = \sum_{j \neq i} a^{0,T}_{ij} - a^{0,C}_{i(n+1)}(1 - (\frac{1-2CF}{1+(n-2)CF})^{m-1}.(\frac{nCF}{1+(n-2)CF})) \quad (5)$$

Since $(1 - (\frac{1-2CF}{1+(n-2)CF})^{m-1}.(\frac{nCF}{1+(n-2)CF})) < 1$ for $m > 1$ and $0 < CF < 1/2$ the theorem follows. $\Delta$.

Although we are unable to eliminate the component $|a^{0,C}_{i(n+1)}|$ completely, Theorem 3 shows that for any $0 < CF < 1/2$ and $m > 1$ the Borda deviation obtained from the revised scores is less than what is obtained from the unrevised scores. Thus, we always do better by revising the scores rather than by not revising it, under a standard and much used method, the Borda Count.

A natural question is, what choice of $CF$ minimize the deviation component and how is the optimal value related to $n$, the number of candidates in the field? Some answers are provided below.

Since $(\frac{1-2CF}{1+(n-2)CF}) < 1$ (and $(\frac{nCF}{1+(n-2)CF})) < 1$), for any $CF < 1/2$, the expression, $y = (\frac{1-2CF}{1+(n-2)CF}).(\frac{nCF}{1+(n-2)CF})$ is maximized - the deviation, $(1 - (\frac{1-2CF}{1+(n-2)CF}).(\frac{nCF}{1+(n-2)CF}))$ is minimized - for $m = 2$. Thus for any $CF < 1/2$, only two iterations are sufficient to achieve the best that can be achieved.

Set $m = 2$. The value of $y$ measures the amount by which the weight of $|a^{0,C}_{i(n+1)}|$ can be reduced below one. Note that $y$ is non-monotone in $CF$ for any given $n$. For each $n$, an optimal value of $CF$ for which $y$ attains a maximum ($(1 - (\frac{1-2CF}{1+(n-2)CF}).(\frac{nCF}{1+(n-2)CF}))$ attains a minimum) can be found numerically. Fig 1 plots $y$ as a function of $CF$, for given values of $n$. It is clear from the expression and the figure that the optimal $CF$ decreases with $n$. That is, as the number of candidates increases, we should use lower values of $CF$ to do the best we can with the revised score, under the Borda method. Note also however, that the optimal $CF$ is strictly positive for finite $n$ however large - implying that it is always advantageous to revise the scores, so long as $n$ is finite.

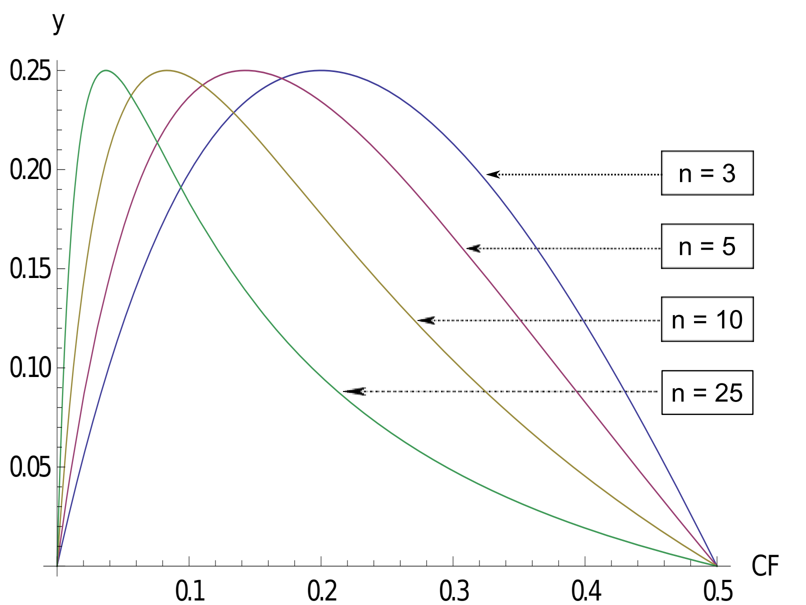


Figure 1: $y = (\frac{1-2CF}{1+(n-2)CF}).(\frac{nCF}{1+(n-2)CF})$, percentage reduction in the Condorcet component for different values of $n$ and $CF$

For an idea about how much reduction in the weight of $|a^{(0,C)}_{i(n+1)}|$ can be acheived, note that some manipulation of the expression $y$ shows that for large but finite $n$ and a choice of $CF = 1/n$, the value of $y$ is close to 0.25. Thus it is possible to reduce the weight of $|a^{(0,C)}_{i(n+1)}|$ by about 25%,[15]for large $n$, irrespective of the magnitude of $|a^{(0,C)}_{i(n+1)}|$. The rule $CF = 1/n$ also provides an approximation for the optimal value of $CF$ for large but finite $n$, without having to solve for it exactly.

Note that sometimes a reduction of 25% may be sufficient. Condorcet components are not always a problem. They are a problem when their weight is large enough relative to the weight contributed by the Basic profiles. Hence a reduction of this weight by 25% may be sufficient to produce consistent results using the Borda method, in many cases. Thus we hope that Theorem 3

[15]The figure of 25% applies to the component $|a^{(0,C)}_{i(n+1)}|$ in the Borda Count and not to the individual pairwise components $|a^{(0,C)}_{ij}|$. For a discussion of how the weights of $|a^{(0,C)}_{ij}|$ are individually affected, see Theorem 4 and discussion below.

is useful from a practical standpoint.

Theorem 3 provides an idea of how effective our method is for the Borda Count. From the point of view of other (pairwise) election methods, it may be of interest to know how the Condorcet terms in the *individual* pairwise score differences behave, under our method. To study this, we define the *residual Condorcet term* for each pair $(i, j)$, after $m$ iterations, as,

$$R_{i,j}(CF,m) = |a_{ij}^{(0,C)}|\frac{p^{(m)}}{t^{(m)}} + |(a_{j(n+1)}^{(0,C)} - a_{i(n+1)}^{(0,C)})|\frac{s^{(m)}}{t^{(m)}}$$

Note that for $CF = 0$, that is if the scores are not revised at all, $R_{i,j}(0) = |a_{ij}^{0,C}|$.

Now assume $m = 2$ and $CF < 1/2$. We ask whether $R_{i,j}(CF,2) < |a_{ij}^{(0,C)}|$ for *each* pair $(i, j)$, for some value of $CF$ and in particular for the optimal value for $CF$ (chosen to minimize the weight of $|a_{i(n+1)}^{(0,C)}|$ in the Borda Count). Unfortunately, $R_{i,j}(CF,2)$ depends on *two* unknown Condorcet components, $|a_{ij}^{(0,C)}|$ and $|(a_{j(n+1)}^{(0,C)} - a_{i(n+1)}^{(0,C)})|$, and a strong answer to this question, analogous to Theorem 3 is difficult to provide. Nevertheless, conditions under which $R_{i,j}(CF,2) < |a_{ij}^{(0,C)}|$ for each pair $(i, j)$, can be provided.

In a slight abuse of terms we describe the term $|a_{ij}^{(0,C)}|$ as the ”known” and the term $|(a_{j(n+1)}^{(0,C)} - a_{i(n+1)}^{(0,C)})|$ as the ”unknown” Condorcet deviations. The descriptors highlight the fact that the first deviation is inherent in the given initial score and remains in it, if we choose not to revise the initial scores. The second deviation is introduced to the revised scores by our method and is thus idiosyncratic. Theorem 4 provides a bound on the relative ratio of ”unknown” to ”known” deviations, $\frac{|(a_{j(n+1)}^{(0,C)} - a_{i(n+1)}^{(0,C)})|}{|a_{ij}^{(0,C)}|}$, which if satisfied for each pair $(i, j)$, an unambiguous answer to our question in the previuos paragraph is possible.

**Theorem 4** $R_{i,j}(CF,2) < |a_{ij}^{(0,C)}|$ *iff* $\frac{|(a_{j(n+1)}^{(0,C)} - a_{i(n+1)}^{(0,C)})|}{|a_{ij}^{(0,C)}|} < n(1 + \frac{1-2CF}{1+(n-2)CF})$ *for all* $(i, j)$.

*Proof*: Follows directly on substitution for $\frac{p^{(m)}}{t^{(m)}}$ and $\frac{s^{(m)}}{t^{(m)}}$

Theorem 4 essentially says that so long as the ”unknown” deviation is no more than a critical multiple (equal to $n(1 + \frac{1-2CF}{1+(n-2)CF})$ and always greater than one) of the ”known” deviation, our method succeeds in reducing the individual Condorcet terms as well. For $CF < 1/2$, the critical multiple, $n(1 + \frac{1-2CF}{1+(n-2)CF})$, is a number between $n$ and $2n$. Thus our method allows the ”unknown” deviation to be at least $n$ times as large as the "known" deviation and still succeeds in reducing the Condorcet terms individually.[16]

### 4.3 An illustrative example from the 2008 US presidential primaries

During the presidential primaries season of May 2008 and shortly before the economy plunged into the worst financial crisis in recent memory (which dramatically altered people's priorities and hence preferences), US voters were excited and deeply divided over three candidates for the US Presidency. Obama and Clinton were two charismatic candidates from traditional "minority" groups on the Democrat side and McCain a popular maverick, was on the Republican side. In varying degrees, all three candidates had great personal appeal for voters cutting across party and political lines. In this section we use our method on published Gallup opinion poll data to try and understand voters' preferences across these three candidates, at that time.

On May, 28, 2008 a Gallup opinion poll had Obama trailing McCain 45% to 46% and Clinton leading McCain, 48% to 44%, amongst *national registered voters* in match-ups. In another opinion poll amongst *Democrat and Democrat leaning voters*, Clinton trailed Obama 44% to 50% in a match-up. Gallup opinion polls conducted a few days before and after this date confirmed the same pattern although the tallies were a little different. No comparable opinion poll data between Clinton and Obama amongst national registered voters was published. However from other evidence and opinion polls conducted during this time, it is safe to assume that amongst independents, Republican and Republican leaning voters, support for Obama was higher than support for Clinton, compared to what it was amongst Democrat and Democrat leaning voters. It may not be far from the truth therefore, to assume (at least as a first cut) that amongst national registered voters, in a hypothetical match-up between Obama and Clinton for presidency, Obama would have lead Clinton by 6%. In course of the analysis below we check whether our qualitative conclusions are sensitive to revising this difference downwards. Clearly, a strong Condorcet component is at play here and we use our method to understand who the voters liked most, once the Condorcet

[16]Note that for $m = 2$, a large and finite $n$, and a choice of $CF = 1/n$, a slight manipulation of the expression yields, $\frac{p^{(2)}}{t^{(2)}} \approx \frac{1}{4}$. This implies that it is possible to reduce the "known" Condorcet term $|a_{ij}^{(0,C)}|$ for each pair $(i, j)$, by about 75%. Thus if $|(a_{j(n+1)}^{(0,C)} - a_{i(n+1)}^{(0,C)})|$ is not significant, the potential gains from revising the scores could be big for the individual pairs.

component is eliminated.[17]

First, note that in all three match ups, an overwhelming majority - 91% in the Obama-McCain race, 92% in the Clinton-McCain race and 94% in the Obama-Clinton race - chose between the two candidates. A small percentage, 3 - 4%, in all three races said they had no opinion; another small percentage, 4 - 5%, said they would vote for neither; and about 1 - 2% said they would vote for a third but unnamed candidate. The preferences of the voters who responded that they had no opinion or that they would vote for neither do not fit the present framework and may be regarded as not having participated in the poll. Accordingly we drop them from the electorate for our analysis. The fact that 1 - 2% of voters said that they would vote for a third but unnamed candidate, indicates that other candidates figured in the voters' preferences. Therefore this situation fits our description of an incomplete field election. However, our framework requires all voters to choose one candidate (their next best) from every pair. Hence, by our definition, these 1-2% of voters have not participated in the election either and are also consequently dropped. Thus assuming that in each race, the voters who actually chose between the two candidates comprised the whole electorate, the pairwise tallies of each candidate in each race are accordingly adjusted upwards to ensure that the electorate in all three races are of the same size. Denoting Obama as 1, Clinton as 2 and McCain as 3, the given normalized pair-wise score differences are seen to be, $a^0_{12} = 0.064$, $a^0_{13} = -0.01$ and $a^0_{23} = 0.0434$.

Using our method, the revised pair-wise scores are $a^1_{12} = 0.039$, $a^1_{13} = 0.042$ and $a^1_{23} = 0.003$. Thus our revised figures has Obama beating McCain by about 4.2%, Clinton beating McCain by a much smaller margin of 0.3% and Obama beating Clinton by about 3.9%. Once the Condorcet component is removed, Obama in fact emerges as the Condorcet winner with clear, almost equal leads over Clinton and Mccain. Clinton is in almost a statistical dead heat against McCain with a very slight lead. The difference between Obama and Clinton is also much less than what the initial pairwise score difference of 0.064 suggests. Finally, to address the issue of national registerd voters vs Democrat and Democrat leaning voters for the Obama-Clinton match-up, note that so long as the initial pairwise score difference between Clinton and Obama amongst national registerd voters is less than 0.064 but greater than 0.025, Obama continues to emerge as the Condorcet winner by

[17] A Borda Count using these initial pairwise scores show that Obama is the Borda winner. Thus more people placed him in the second position compared to Clinton or McCain. Voters who had Clinton as their number 1 preferred McCain relatively more than Obama. Voters who had McCain as their number 1 preferred Obama more than Clinton.

our method.

## References

ARROW, K. J. (1963): *Social Choice and Individual Values*. Wiley.

BALINSKI, M., A. JENNINGS, AND R. LARAKI (2009): "Monotonic Incompatibility Between Electing and Ranking," *Economics Letters*, 105.

BALINSKI, M., AND R. LARAKI (2007): "A Theory of Measuring, Electing and Ranking," *Proceedings of the National Academy of Sciences*.

BLACK, D. (1958): *The Theory of Committees and Elections*. Cambridge.

BORDA, J. C. (1781): *Memoire Sur Les Elections Au Scrutin.* Histoire de L'Academie Royale des Sciences.

CHANDRA, A., P. KARRA, AND M. DOROTHY (2008): "Implications of Arrow's Theorem in Modeling of Multiscale Phenomena: An Engineering Approach," NSF-DMMI Grantee Conference, Knoxville, TN, 1/7-10,2008.

CONDORCET, M. (1785): *Essai Sur L'application De L'analyse A La Probabilite Des Decisions Rendues A La Pluralite Des Voix*. Paris.

COPELAND, A. H. (1951): "A Reasonable Social Welfare Function," Mimeo.

HODGE, J. K., AND R. E. KLIMA (2005): *The Mathematics of Voting and Elections: A Hands-On Approach*. American Mathematical Society.

KEMENY, J. (1959): "Mathematics Without Numbers," *Daedalus*, 88, 571–591.

MOULIN, H. (2003): *Fair Division and Collective Welfare*. MIT Press.

RASCH, B. E. (1995): "Parliamentary Voting Procedures," in *Parliaments and Majority Rule in Western Europe*, ed. by H. Doring. Campus Verlag.

SAARI, D. G. (1999): “Explaining All Three Voting Outcomes,” *Journal of Economic Theory*, 87, 313–355.

——— (2000a): “Mathematical Structure of Voting Paradoxes I. Pairwise Votes,” *Economic Theory*, 15, 1–53.

——— (2000b): “Mathematical Structure of Voting Paradoxes II. Positional Methods,” *Economic Theory*, 15, 55–102.

SLATER, P. (1951): “Inconsistencies in a Schedule of Paired Comparisons,” *Biometrica*, 48, 303–312.