The Source of Election Results:

An Empirical Analysis of Statistical Models of Voter Behavior

T. Nicolaus Tideman

Department of Economics, Virginia Polytechnic Institute and State University

Blacksburg, VA 24061

ntideman@vt.edu

Florenz Plassmann*

Department of Economics, State University of New York at Binghamton

Binghamton, NY 13902-6000

fplass@binghamton.edu

This version: June 29, 2008

Abstract:

We consider six models of voter behavior that might govern the statistical process of vote-casting. From an empirical analysis of "election-like" data we conclude that a spatial model of voting, augmented by a specified error structure, describes this process much better than the other five models. An important benefit of identifying the properties of this process is that doing so permits one to evaluate voting methods in terms of their success in identifying the proper winners of elections, so that it is not necessary to rely entirely on comparisons of logical properties in evaluating voting methods.

Journal of Economic Literature Classification Codes: C4, D72

Keywords: Borda, Condorcet, Estimated Centrality, voter behavior, Pólya distribution

* Corresponding author. Phone: (607) 777-4304, Fax: (607) 777-2572.

1. INTRODUCTION

What method of determining the winner is best for processing elections with three or more candidates? Voting theorists customarily evaluate competing voting methods by comparing their logical properties.[1] Because no method possesses all properties that are desired, and there is no consensus on which properties are most important, there is also no consensus on which voting method is best. Voting theorists have also used theoretical analyses as well as Monte Carlo simulations to determine the frequency with which events such as the selection of the Condorcet winner, if there is one, occur under different voting methods.[2] Because this approach requires a model of voter behavior that describes how voters cast their ballots, and there is no consensus on which model of voter behavior describes vote casting best, this strategy has also not led to a consensus on which voting method is best.

We argue that any inquiry into which voting method is most attractive requires that one first identify the statistical process that governs vote-casting through an empirical analysis of actual elections. This makes it possible both to obtain more relevant estimates of the frequency of such voting events as the election of Condorcet winners, and also to construct simulated elections in a way that permits the right winner to be known, so that one can estimate the frequency with which different voting methods select the right winner. In this paper, we motivate and investigate several statistical models and identify a promising model of voter behavior on the basis of an empirical analysis of election-like data. We use our results to undertake a preliminary analysis of three voting methods, but defer more detailed analyses of voting methods to future research.

[1] See, for example, Saari (1999) and Tideman (2006).

[2] See, for example, Chamberlin and Cohen (1978) and Merrill (1984) for Monte Carlo analyses and Cervone *et al*. (2005) for a theoretical analysis of this question.

The intuition behind our proposed strategy is explained best by the observation that the task of finding the most attractive voting method can be viewed as being analogous to the task of determining the best estimator of a statistical parameter, such as the parameter μ that represents the mean of a normal distribution, in an analysis of observational data. In the case of voting, the parameter to be estimated is the winner of the election. Theoretical analyses that concentrate on the logical properties of voting methods are thus comparable to the derivation of the theoretical properties of different statistical estimators. Similarly, Monte Carlo estimates of the frequencies with which certain events occur under different voting methods are comparable to Monte Carlo evaluations of estimators. However, in the same way in which different estimators may each be optimal for analyses of data from different statistical models, different voting methods may each be optimal for voting data that are derived from different statistical models of voter behavior. Thus it is not surprising that theoretical analyses of voting have been unable to identify an objectively best voting method, since these analyses have not been linked to empirical inquiries into statistical properties of vote-casting.[3]

The first step in the analysis of observational data is to determine the statistical model that is most likely to have generated the available data (the "data generating mechanism").[4] Thus we need to specify first the statistical model of voter behavior that describes observed rankings of candidates that voters supply as a function of the circumstances in a specific election, and the frequencies of different electoral circumstances. Only then is it meaningful to estimate and analyze the frequencies with which different voting methods reach certain results, for example,

---

[3] Beginning with Condorcet (1785), many voting theorists have analyzed voting methods as maximum likelihood estimators (see, for example, Young, 1986, Drissi-Bakhkhat and Truchon, 2004, Conitzer and Sandholm, 2005, and Truchon, 2006). However, these authors did not evaluate the estimators on the basis of actual election data and corresponding statistical models of voter behavior.

[4] See, for example, Hendry (1980) and Spanos (1986).

identify the correct winner or choose a Condorcet winner. Our surprisingly strong results indicate that this line of inquiry is worth pursuing.

It is worth emphasizing how our strategy differs from previous empirical inquiries into voting outcomes. By focusing on statistical models that describe voter behavior in general, our strategy differs fundamentally from analyses that investigate the properties of voting methods in individual elections (for example, Chamberlin *et al*.,1984, Regenwetter and Grofman, 1998a, Brams and Fishburn, 2001, Saari, 2001, Regenwetter *et al*., 2002, Tsetlin and Regenwetter, 2003, and Regenwetter and Tsetlin, 2004). In contrast to Monte Carlo simulations that postulate certain models of voter behavior to generate the data (for example, Campbell and Tullock, 1965, Chamberlin and Cohen, 1978, Merrill, 1984, Berg, 1985, and Adams, 1999), we use data from actual elections to evaluate the adequacy of different models of voter behavior. Finally, our primary goal is not to create a statistical model that is capable of generating artificial data that look like election data (as has been done by Chamberlin and Featherston, 1986), but rather to develop a statistical framework that permits us to evaluate different plausible models of voter behavior on the basis of observed election data.

We start our investigation from the two premises that, first, there is a correct or best outcome and second, the purpose of voting is to identify this outcome. Neither premise is self-evident. Voting can certainly be meaningful even in the absence of a correct outcome if it helps voters make collective decisions in ways that they find satisfactory. It is also not obvious what “the correct outcome” means. If consensus on the correct outcome could be reached, then there would be no need to vote. Still, voters might agree on the properties of the correct outcome but not agree on which outcome this is. For example, voters might agree that the outcome is best that leads to the lowest collective utility loss but not agree on which outcome achieves this. In such a case the outcome does not need to be objectively correct; it suffices that the voters agree on the criterion that defines the best outcome. (We treat the terms "correct outcome" and "best outcome" as synonyms.) The notion that there exists a correct outcome has been part of the analysis of voting for a long time; it is an essential component of the literature on jury theorems that Condorcet started with his 1785 *Essai*, and we view our analysis as a part of this tradition.

We consider six models of voter behavior to be interesting candidates for our initial inquiry. First, it may simply be incorrect to assert that in any given election, some rankings of candidates are systematically more likely to occur than others. The corresponding model of voter behavior assumes that in every election, all rankings are equally likely to be reported by every voter. This assumption is known in the voting literature as the "impartial culture assumption." The impartial culture assumption implies that voting results contain no information that is useful for identifying the correct winner. Analyzing election data with this model therefore provides information about whether our proposed strategy of inquiry has any merit. If there is evidence that voters discriminate among rankings in systematic ways and that we can therefore reject the model of equally likely rankings, then it is meaningful to ask whether their voting behavior makes any of the known voting methods more likely than others to identify the correct winner.

Our second model of voter behavior assumes that there is a single vector of unequal probabilities for different rankings, which is valid for all elections. This model, supplemented by an error process, was proposed by Chamberlin and Featherston (1986).

Our third model of voter behavior assumes that every vector of probabilities for the possible strict rankings of the candidates (with probabilities summing to one) is just as likely as every other such vector. This assumption is known in the voting literature as the "impartial anonymous culture assumption" (see Kuga and Nagatani, 1974, and Gehrlein and Fishburn, 1976), and voting theorists frequently use it to calculate the likelihood that certain voting events will occur. For example, Saari (1990) uses this assumption to analyze the likelihood of strategic voting under different voting methods, Gehrlein (2002) uses it to analyze the likelihood of observing Condorcet's paradox, and Cervone *et al*. (2005) use it to analyze the likelihood that a Condorcet candidate, if it exists, will win the election. Although voting theorists generally emphasize that they do not necessarily believe that equally likely probabilities of strict rankings describe voter behavior better than any other model does, the frequent use of this assumption suggests that it is informative to test whether this model is at all likely to describe actual voter behavior.

The remaining three models are our main models of interest. The fourth and fifth models are inspired by the Borda voting method and the Condorcet-Kemeny-Young voting method (henceforth "Condorcet method"), which are the voting methods over which the academic dispute about the relative merit of voting methods started.[5] We consider two models of voter behavior for which these voting methods are the respective maximum likelihood estimators. We are not aware of any previous study that analyzes voting methods with either of these two models. Our sixth and final model of voter behavior is based on the widely-used spatial model of voting (see Enelow and Hinich, 1984 and 1990). Chamberlin (1978), Merrill (1984), and Adams (1997) have used spatial models of voter behavior for Monte Carlo analyses of Condorcet's paradox, and concluded that the results of their analyses differ considerably from the results that they obtained under the impartial culture assumption. We use a somewhat more general spatial model of voter behavior that provides the motivation for Good and Tideman's (1976) voting method, which Tideman (2006) calls the "Estimated Centrality" method.

[5] See Black (1958, pp. 156 – 80) as well as the recent exchange between Risse (2001, 2005) and Saari (2003, 2006).

In the remainder of this paper, we assess the validity of these six models of voter behavior, using data from 913 "elections" that we construct from the "thermometer scores" that are part of the surveys conducted by the American National Election Studies (ANES). In Sections 2 and 3, we develop the statistical framework and the six models of voter behavior, and explain our model assessment strategy. We describe our data and report the results of our statistical analysis in Section 4. We find strong evidence supporting the hypothesis that the combination of the spatial model of voter behavior and a specified error structure that adds a small amount of variation to the probabilities specified by the spatial model provides a better explanation of voting patterns than any of the other models.

In Section 5, we use the spatial model of voter behavior, augmented by the specified error structure, to simulate artificial elections, and we assess the frequencies with which three voting methods—the Borda method, the Condorcet method, and the Estimated Centrality method—determine the correct winner. We find that the accuracy of identifying the correct winner depends on the number of voters far more than it depends on the voting method. However, for any given number of voters there are significant differences in the accuracy of these voting methods, so it is valuable to take account of accuracy when choosing a voting method. We find that the Borda method has the highest rate of accuracy if the number of voters is small, while the Estimated Centrality method is most accurate when the number of voters is large.

An important advantage of using survey data rather than actual election data to assess the validity of models of voter behavior is that survey respondents are less likely to have any reason to report rankings that differ in systematic ways from their true rankings. Thus our data set is likely to provide us with information about voter behavior that is little affected by strategic considerations. We emphasize that we follow the standard practice in empirical analyses of voting methods and do not evaluate any voting method according to its resistance to strategizing. The incorporation of estimates of resistance to strategizing into the analysis of voting methods may lead to different conclusions about the relative attractiveness of voting methods. Saari (1990) proposes ways of assessing a voting method's theoretical susceptibility to strategizing, but he undertakes his analysis under the assumption that the impartial anonymous culture prevails. Our analysis indicates that assessing a voting method's resistance to strategizing on the basis of a model of voter behavior developed from actual election data is likely to provide further insights.

2. A Statistical Model of Vote-Casting

Consider an election with M candidates, in which N voters are asked to each submit a ranking of the candidates. There are $M!$ possible strict rankings. If p_r is the probability that a voter submits ranking r, $r = 1, \ldots, M!$, and $\Sigma p_r = 1$, then we can view the number of votes for ranking r, N_r, as a random variable whose statistical model describes the probabilities of the different possible outcomes of the vote-casting process. This statistical model can be specified through the joint distribution of the N_rs. Assuming that the p_rs are deterministic, that they are the same for all voters, and that all voters submit their rankings independently, the N_rs follow a multinomial distribution with density function[6]

$$f(N_1, \ldots, N_{M!}; N, p_1, \ldots, p_{M!}) = \prod_{r=1}^{M!} p_r^{N_r} \frac{N!}{\prod_{r=1}^{M!} N_r!}, \tag{1}$$

whose first two moments are $E[N_r] = Np_r$, $\mathrm{Var}[N_r] = Np_r(1 - p_r)$, and $\mathrm{Cov}[N_r, N_s] = -Np_rp_s$.

Note that we model the outcome of the vote-casting process in terms of the probabilities with which a voter submits any of the possible rankings, and not in terms of the probabilities that

[6] Generally, the multinomial distribution describes the probabilistic structure of any series of independent and identical Bernoulli trials with constant probabilities and multiple possible outcomes.

a voter ranks a particular pair of candidates one way or the other. The latter has been the standard framework for evaluating voting methods as maximum likelihood estimators since Condorcet's *Essai*. A frequently made assumption in this framework is that these probabilities are independent across pairs of candidates, which permits voters to have non-transitive preferences.[7] In contrast, modeling the probabilities of submitting different rankings of the candidates imposes transitivity on the voters' preferences naturally, and is therefore more likely to reflect the actual statistical model that governs the outcome of the vote-casting process.

Now assume that there is a correct ranking r^* and that the highest ranked candidate in r^*, candidate m^*, is the correct winner. We incorporate the notions of a correct ranking and correct winner into the statistical model of the N_rs through a model of voter behavior that describes the probabilities with which voters chose any of the rankings. We consider six models of voter behavior, which we motivate and describe in the following section.

3. SIX MODELS OF VOTER BEHAVIOR

3.1. *Equally likely rankings ("impartial culture")*

The simplest model of voter behavior assumes that all rankings are equally likely, or $p_r = 1/M!$ (the "impartial culture assumption"). Voting theorists generally acknowledge that they consider this model to be unrealistic, and Tsetlin *et al.* (2003) show that the impartial culture assumption maximizes the likelihood of observing Condorcet's paradox for three candidates. Nevertheless, if this assumption is correct, then voting is at best a useful mechanism for making a collective decision but not a useful mechanism for identifying the correct outcome, and any attempt to determine the voting method most likely to identify the correct winner will be futile. Thus the

[7] See Young (1995, p. 55) and Drissi-Bakhkhat and Truchon (2004, p. 167).

model of equally likely rankings represents a benchmark against which we measure the value of voting in the other five models.

3.2. *Unequally likely rankings, with probabilities that are stable across elections*

Instead of assuming that all rankings are equally likely, one can assume that in each election, each candidate occupies a specifiable ranking niche (first, second, etc.), and that for each possible ranking of the candidates described by this niche, there is a probability, the same across elections, that this ranking will be used. Chamberlin and Featherston (1986) use a variation on this model to analyze data from five presidential elections of the American Psychological Association. They provide evidence that this model of voter behavior explains actual election data better than the model of equally likely rankings.

3.3. *Equally likely vectors of probabilities for rankings ("impartial anonymous culture")*

The next simplest model of voter behavior assumes that every vector of probabilities for rankings, $(p_1, \ldots, p_{M!})$, such that the probabilities sum to 1, is as likely to prevail as every other such vector (the "impartial anonymous culture assumption"). We are not aware of any formal test of the adequacy of this frequently-made assumption, which makes it interesting to test whether the impartial anonymous culture assumption describes actual voter behavior better than any of the other models that we analyze.

3.4. *A model of voter behavior inspired by the Borda method*

Borda asserted that the amount of evidence in favor of the proposition that a particular candidate was best increased by the same amount each time a voter raised a candidate one notch in his ranking.[8] Thus when a voter ranks a candidate in position k, $k = 1, \ldots, M$, the Borda method assigns $M - k$ points to this candidate. Candidate m's Borda score is the sum of the points that m obtains from all voters, and the candidate with the highest Borda score is the Borda winner.

To derive a model of voter behavior that justifies the Borda method, we define the probabilities of the rankings, p_r, $r = 1, \ldots, M!$, in such a way that the Borda winner is a maximum likelihood estimate of the correct candidate. Consider any two rankings r and s in which the correct candidate m^* is ranked in positions k_r and k_s respectively. If a voter submits ranking r rather than ranking s, then the difference in the points that the Borda method assigns to candidate m^* is $k_r - k_s$. Recall our assumptions that the p_rs are the same for all voters and that the N voters submit their rankings independently. For the Borda method to be a maximum likelihood estimator of m^*, the difference in the log of the likelihood function from submitting ranking r rather than s must therefore be proportional to the difference in points that the Borda method assigns to m^* under the two rankings, or

$$\ln p_r - \ln p_s = \alpha (k_r - k_s) \tag{2}$$

for some constant α. This implies that

$$p_r = c_1 e^{\alpha k_r} \tag{3}$$

with

$$c_1 = 1 / \left((M-1)! \sum_{m=0}^{M-1} e^{m\alpha} \right) \tag{4}$$

to ensure $\Sigma p_r = 1$. The assumption that the p_rs are the same for all voters implies that α is the same for all voters. To confirm that such probabilities imply that the candidate with the greatest Borda score is most likely to be the correct winner, denote the ranking supplied by voter n by

[8] Black (1958), pp. 157-158.

$r(n)$, and compute the Borda likelihood function LH_B of the N rankings submitted by the N voters as

$$LH_B \propto \prod_{n=1}^{N} p_{r(n)} = \left(c_1^N\right) e^{\alpha \sum_{n=1}^{N} k_{r(n)}} \tag{5}$$

where $\Sigma k_{r(n)}$ is candidate m^*'s Borda score. The candidate with the highest Borda score is therefore the maximum likelihood estimate of the correct candidate m^* if the Borda model of voter behavior defines the probabilities of the $M!$ rankings.

Table 1 shows the p_rs required by the Borda model of voter behavior, for an election with three candidates A, B, and C and a given value of α. The entry in each row of the final column is the probability of the rankings in the three preceding columns, conditional on the correct candidate being the one listed in the respective column heading. Because the Borda method assigns scores to candidates according to their positions within a ranking, it assigns the same probability to all rankings that rank the correct candidate in the same position.

Table 1. Probabilities of rankings for the statistical model inspired by the Borda method

Correct candidate			
A	B	C	
Ranking	Ranking	Ranking	Probability
ABC, ACB	BAC, BCA	CAB, CBA	$p_r = c\ e^{2\alpha}$
BAC, CAB	ABC, CBA	ACB, BCA	$p_r = c\ e^{\alpha}$
BCA, CBA	CAB, ACB	BAC, ABC	$p_r = c$

We calibrate the Borda model to observed election data by applying the Borda voting method. In our empirical analysis in Section 4, we assume that α follows a Gamma distribution with two parameters. The evaluation of a single election provides an estimate of α as the value that maximizes the likelihood of the election data. Application of the Borda method to multiple elections yields multiple estimates of α, which we use to calibrate the parameters of the distribution of α.

3.5. *A model of voter behavior inspired by the Condorcet method*

While the Borda method assigns a score to each of the M candidates, the Condorcet method assigns a score to each of the $M!$ possible rankings. Condorcet argued that the amount of evidence in favor of the proposition that a particular ranking of the candidates is the correct ranking is the sum over all ballots of the number of pairs of candidates that the proposed ranking and the ballot place in the same order.[9] Thus to assign a score to ranking r, the Condorcet method counts, for each ballot, the number of pairs of candidates that are ranked on the ballot in the same way as in ranking r. The sum over all N ballots of the number of pairs ranked the same on the ballot as in r is ranking r's Condorcet score. The ranking with the highest Condorcet score is the winning ranking by the Condorcet method, and the winning candidate is the candidate at the top of the winning ranking. Condorcet's explanation of his method was opaque and contained errors; Kemeny proposed the same voting method in the twentieth century, and Young explained (to our satisfaction at least) how Condorcet's intention could be understood despite his errors.[10]

---

[9] See Young (1988).

[10] See Kemeny (1959) and Young (1988).

To derive a model of voter behavior that justifies the Condorcet method, we define the probabilities of the rankings p_r, $r = 1, \ldots, M!$, in such a way that the Condorcet method is a maximum likelihood estimator of the correct candidate. Define n_{sr} as the number of pairs of candidates that are ranked the same in ranking s as in ranking r. Consider any two ballots whose rankings, r and s, agree with the correct ranking, r^*, on n_{rr*} and n_{sr*} pairs respectively. If a voter submits a ballot with ranking r rather than s, then this voter's contribution to ranking r^*'s Condorcet score changes by $n_{rr*} - n_{sr*}$. For the Condorcet method to be a maximum likelihood estimator of r^*, the difference in the log of the likelihood function from submitting ranking r rather than s must be proportional to the difference between n_{rr*} and n_{sr*}, or

$$\ln p_r - \ln p_s = \alpha (n_{rr*} - n_{sr*}) \tag{6}$$

for some constant α. This implies that

$$p_r = c_2 \, e^{\alpha n_{rr*}} \tag{7}$$

with

$$c_2 = 1 \Big/ \left(\sum_{m=0}^{\frac{(M-1)M}{2}} f(m, M) e^{m\alpha} \right) \tag{8}$$

to ensure $\Sigma p_r = 1$, where $f(m, M)$ is the frequency distribution of Kendall's τ.[11] Thus if we denote the ranking on voter n's ballot by $r(n)$, then the Condorcet likelihood function LH_C of the N rankings submitted by the N voters is

$$LH_C \propto \prod_{n=1}^{N} p_{r(n)} = \left(c_2^N\right) e^{\left(\alpha \sum_{n=1}^{N} n_{r(n)r^*}\right)} \tag{9}$$

[11] See Kendall and Gibbons (1990, pp. 91 – 92)

where $\Sigma n_{r(n)r*}$ is ranking r^*'s Condorcet score. The ranking with the highest Condorcet score is therefore the maximum likelihood estimate of the correct ranking r^* if the Condorcet model of voter behavior defines the probabilities of the M! rankings.

Table 2 shows the p_rs of the Condorcet model of voter behavior for an election with three candidates A, B, and C. As before, the entry in each row of the final column is the probability of the rankings in the preceding columns, conditional on the ranking with the greatest probability of being selected by voters being the one listed in the respective column heading. As with the Borda model, we assume that α follows a gamma distribution with two parameters, which we calibrate using data from multiple elections.

Table 2. Probabilities of rankings for the statistical model inspired by the Condorcet method

Correct ranking						
ABC	ACB	CAB	CBA	BCA	BAC	
Ranking	Ranking	Ranking	Ranking	Ranking	Ranking	Probability
ABC	ACB	CAB	CBA	BCA	BAC	$p_r = c\ e^{3\alpha}$
BAC, ACB	ABC, CAB	ACB, CBA	CAB, BCA	CBA, BAC	BAC, ABC	$p_r = c\ e^{2\alpha}$
BCA, CAB	BAC, CBA	ABC, BCA	ACB, BAC	CAB, ABC	CBA. ACB	$p_r = c\ e^{\alpha}$
CBA	BCA	BAC	ABC	ACB	CAB	$p_r = c$

3.6. *A spatial model of voter behavior*

Assume that voters care about the "attributes" of candidates. These attributes form a multi-dimensional "attribute space." Every voter has an indifference map in attribute space, which contains an "ideal point" that describes the quantities of each attribute that the voter's ideal candidate would possess. Actual candidates also possess specifiable quantities of each attribute and therefore have locations in attribute space. We assume that voters agree on the locations of the actual candidates. If attribute space has at least $M - 1$ dimensions and the candidates are in "general position," where a slight change in the position of any one candidate does not change the dimensionality of the space that they span, then the positions of the M candidates in attribute space span an $M - 1$ dimensional "candidate space" that is a subspace of attribute space. Voters' indifference maps are defined in candidate space through their definition in attribute space.

To complete the model, we need to specify the distribution of the voters' ideal points and the shapes of the voters' indifference maps. We follow Good and Tideman (1976) and assume that the positions of voters' ideal points in attribute space follow a spherical multivariate normal distribution and that these positions are independent of each other, which implies that the distribution of "relative" ideal points in candidate space is spherical multivariate normal as well.[12] We further assume that every voter's utility loss from the choice of a particular candidate is the same increasing function of the distance between the candidate's location in candidate space and the voter's relative ideal point in candidate space, so that every voter's indifference surfaces are concentric spheres centered on the voter's ideal point. None of these assumptions is conceptually necessary and each could be replaced—at a cost of more complex calculations—if there is evidence that is does not represent observed voting data sufficiently well.[13]

We define the correct winner m^* as the candidate whose election leads to the lowest collective utility loss among the voters, and we assume that the collective utility loss is the sum of the losses of the individual voters. In conjunction with the assumptions about the voters' ideal points and preferences, this implies that the mode of the distribution of ideal points in candidate

[12] A copy of Good and Tideman (1976) is available at http://bingweb.binghamton.edu/~fplass/GoodTideman.pdf.

[13] See Good and Tideman (1976) for a discussion.

space represents the socially most attractive location for a candidate and that the best candidate among those who are available is the one who is closest to the mode.

Now assume that there is a set of candidates for which every voter submits a truthful ranking that reflects his ideal point, his indifference surfaces, and the positions of the candidates. The expected share of votes that each ranking receives is the integral of the density of voter ideal points over the region of the candidate space in which voters choose this ranking. Because it is straightforward to illustrate the computation of vote shares graphically for the case of three candidates and because we will restrict our empirical analysis to elections with three candidates, we describe this computation for $M = 3$, where the candidates' positions in attribute space span a two-dimensional "candidate plane."[14] In this case, our assumptions about the distribution of voters' ideal points and the loss of utility in terms of the distance from a voter's ideal point to a candidate's location imply that the relative ideal points in the candidate plane have a circularly symmetric bivariate normal distribution, and that every voter's indifference contours are concentric circles around his ideal point.[15]

To determine how voters rank the three candidates, consider the triangle in the candidate plane that is formed by the locations of the three candidates, A, B, and C. We divide the candidate plane into six sectors by drawing the perpendicular bisectors of the three sides of this triangle. These bisectors intersect at the triangle's circumcenter, *P*. For the voters' ideal points in each sector, the distances to the locations of the three candidates have a unique rank order. These rank orders are indicated in Figure 1, together with the mode of the circular bivariate normal distribution at *O*. The integral of the density function of this distribution over each sector is the

[14] The case when all candidates' attributes lie in a single line requires special treatment because not all of the 6 possible rankings of the candidates occur, but it does not pose conceptual difficulties. See Good and Tideman (1976, pp. 380 – 381) for a description of the general case with $M > 3$.

[15] See Good and Tideman (1976, p. 374).

expected value of the fraction of the voters who rank the candidates in the order corresponding to the sector's rank order. These six integrals determine the probabilities p_r, $r = 1, \ldots, 6$ of the six rankings in the statistical model of the outcome of the vote casting process with 3 candidates. Note that even though sectors that are opposite each other have the same angle, they do not have the same integral of the density function (and therefore do not imply the same p_r), unless O is not inside either of the sectors and the two lines that form the sectors come equally close to O.

The rank order of the sector that contains the mode of the bivariate normal distribution describes the correct ranking r^*, and the candidate whose election leads to the lowest collective

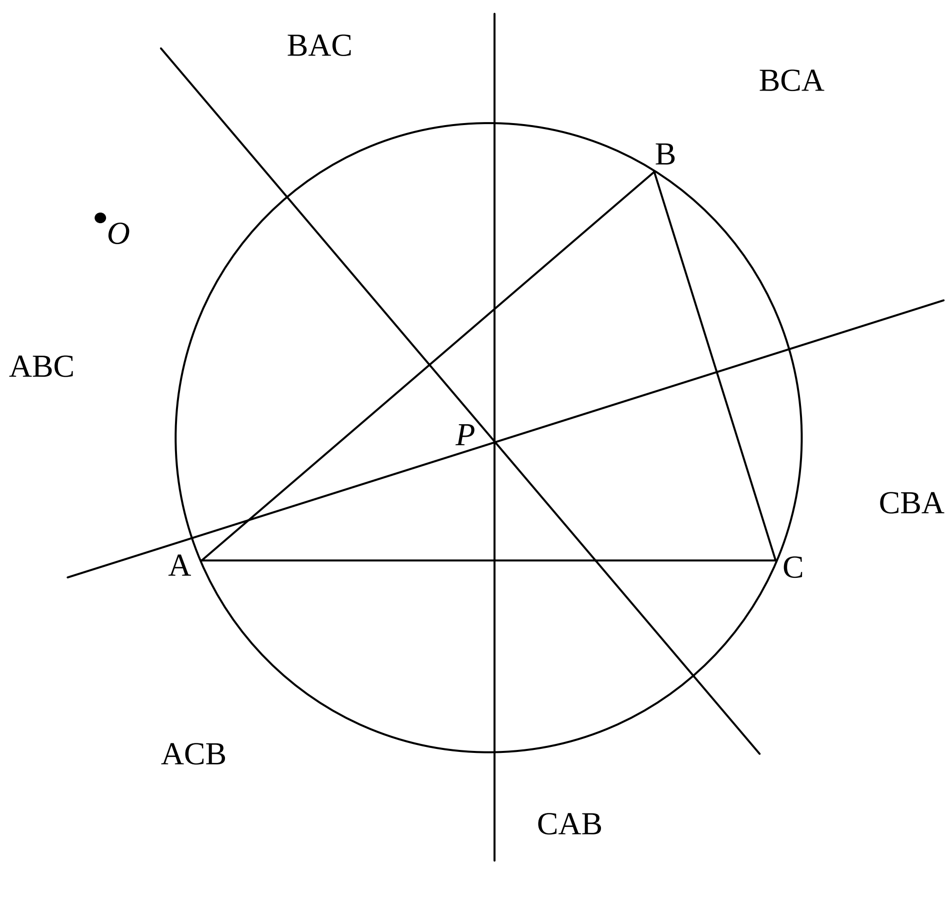

Figure 1. Division of the candidate plane into six sectors by drawing the perpendicular bisectors of the three sides of the triangle formed by the candidates' locations, and the associated rank orders of the sectors. (The figure is taken from Good and Tideman, 1976, p. 372.)

utility loss among the voters (the correct winner, m^*) is the highest ranked candidate in r^*. Thus the Estimated Centrality method, using observable vote shares of the six rankings to identify the sector that contains the mode of the bivariate normal distribution, is a voting method inspired by the spatial model. To implement the spatial model one must place the borders between pairs of adjacent rankings in such a way as to create sectors that match the six probabilities p_r as closely as possible to the six observed vote shares, q_r. Because the construction of the relative locations from the observed q_rs has only four degrees of freedom while the p_rs have five degrees of freedom, a perfect match is generally not possible.[16] Figure 2 shows one way of using the four degrees of freedom. The intersection of the perpendicular bisectors P is placed at the origin of a

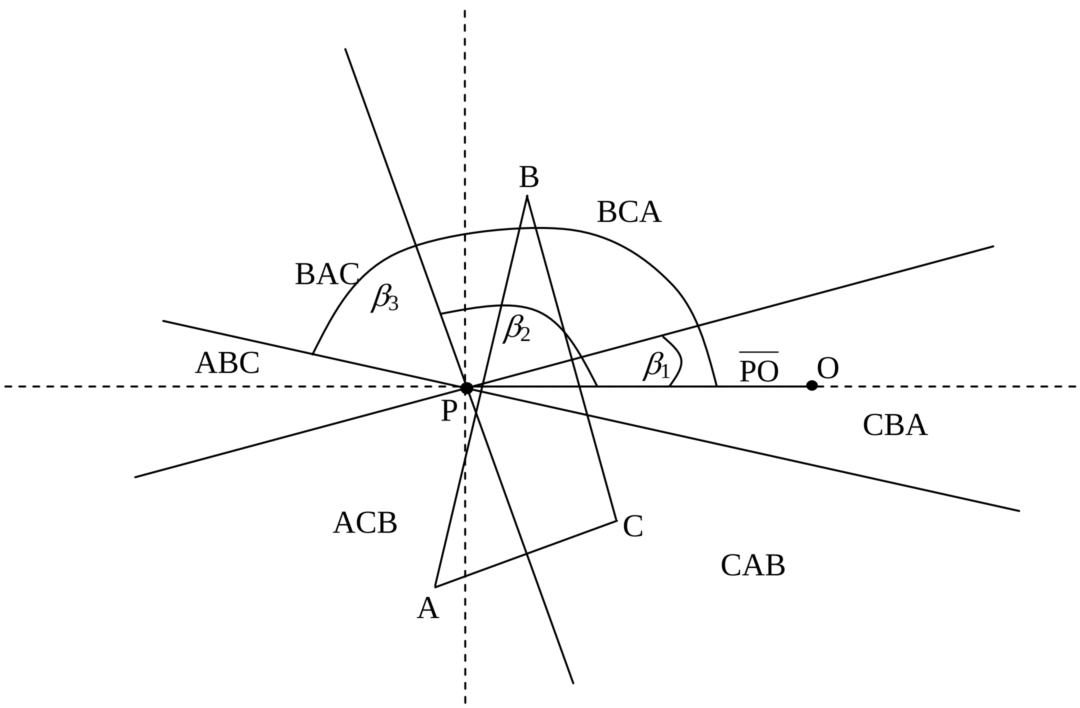

Figure 2. The four parameters $\overline{PO}$, β_1, β_2, and β_3 that define the spatial model.

[16] The vote shares p_rs, which are defined relative to a given set of candidates' locations, have five rather than six degrees of freedom because they must sum to 1. The relative locations of the candidates that one can derive from the q_rs have only four degrees of freedom because they are independent of rotations around O and also independent of changes that move the locations of all candidates proportionately along rays emanating from the triangle's circumcenter.

Cartesian coordinate system. The fact that the vote shares are independent of rotations around the mode of the distribution of voters' ideal points, O, permits us to rotate the coordinate system so that O is located on its horizontal axis. The first degree of freedom then specifies the distance between P and O. The remaining degrees of freedom specify the angles β_1, β_2, and β_3 formed by the line $\overline{PO}$ and the three perpendicular bisectors. Thus any feasible set of values of the four degrees of freedom corresponds to a set of p_rs.

We find the parameters that provide the best match of the p_rs to the q_rs by identifying the angles of the perpendicular bisectors and the location of the mode of the voters' ideal points that maximize the model's log-likelihood ratio (LLR)

$$LLR = \ln\left(\prod_{r=1}^{M!} (p_r)^{N_r} \Bigg/ \prod_{r=1}^{M!} (q_r)^{N_r} \right), \tag{10}$$

that is, the logarithm of the likelihood of the predicted shares p_r, divided by the likelihood of the observed shares q_r, where N_r is the observed number of votes for ranking r. Alternatively, we could minimize the sum of the squared residual differences between the p_rs and q_rs, weighted by the inverses of their estimated variances,

$$WSSR = \sum_{r=1}^{6} \frac{(p_r - q_r)^2}{Np_r(1-p_r)}. \tag{11}$$

Our simulations suggest that the two measures yield approximately equally good matches of the p_rs to the q_rs, and we decided to use the LLR in our calculations.

The computation of the p_rs associated with a proposed set of parameters requires the numerical computation of $M!$ areas under an $(M-1)$-variate normal distribution. Currently we are able to compute such integrals only for bivariate normal distributions, and we must therefore restrict our analysis to three-candidate elections.[17] We use the algorithm described in DiDonato and Hageman (1980) to compute the integral of the bivariate normal distribution over each sector. Because it is straightforward in theory to apply the Estimated Centrality method to elections with more than three candidates, it should be possible to extend our analysis in that direction.

We calibrate the spatial model to observed election data by applying the Estimated Centrality voting method. The evaluation of a single election provides one set of values for the angles of the bisectors and the location of the mode, and the evaluation of multiple elections provides information about the distributions of these parameters. In our Monte Carlo analyses in Sections 4 and 5, we assume that the three angles between pairs of bisectors follow a tri-variate Dirichlet distribution with three parameters (two of the three shares of a semi-circle and a variance parameter), and that the distance between P and O follows a Weibull distribution with two parameters.

3.7 *Model assessment*

We use data from multiple elections to determine which among models 1, 2, 4, 5, and 6 provides the statistically best description of voter behavior. Model 3 (the impartial anonymous culture) requires a different form of assessment (see Section 4.2.2) because the model does not rule out any election that might be observed. For each of the other five models, we compute three measures of fit. The first two measures assess how well the model explains observed election

[17] It is common practice in the literature on probabilistic evaluations of voting methods to restrict the number of candidates for technical reasons. See, for example, Regenwetter and Grofman (1998b), Gehrlein (2002), and Tselin *et. al.* (2003).

data. The third measure assesses the degree to which the model can simulate data that are statistically similar to the observed election data.

We assess how well a model explains each of the observed elections by calibrating its unknown parameters to the observed elections, and then compare the predicted vote shares with the observed vote shares. We assess each model's fit through its mean *LLR*,

$$\text{mean } LLR = \frac{1}{E}\sum_{e=1}^{E}\ln\left(\prod_{r=1}^{M!}(p_{re})^{N_{re}} \Big/ \prod_{r=1}^{M!}(q_{re})^{N_{re}}\right), \tag{12}$$

that is, the mean over the E observed elections of the logarithm of the likelihood of the M! shares p_{re} predicted by the best parameterization of the respective model, divided by the likelihood of the observed shares q_{re}. The algebraically greater a model's mean *LLR*, the better is the model's fit.

As an alternative measure of fit, we determine each model's mean weighted sum of squared residuals,

$$\text{mean } WSSR = \frac{1}{E}\sum_{e=1}^{E}\left(\sum_{r=1}^{M!}\frac{(p_{re}-q_{re})^2}{N_e p_{re}(1-p_{re})}\right), \tag{13}$$

that is, the mean over the E observed elections of the sum of squared differences between each of the M! observed vote shares, q_r, and the corresponding predicted vote shares, p_r, weighted by the inverse of the estimated variance in the number of voters. The lower a model's mean *WSSR*, the closer on average are the model's predicted vote shares to the actual vote shares.

An alternative way of assessing a model's adequacy is to ask whether the model is able to simulate artificial election data that yield the same distribution of *LLR*s or *WSSR*s as observed election data. If the distribution in actual elections differs from the distribution in simulated elections, then this is evidence against the model.

A formal way of assessing the degree to which different models depart from actual elections is to introduce an "error function" in the form of an assumption that the p_rs for each election are random variables rather than being specified directly by a model of voter behavior. An intuitive assumption is that the p_rs follow a Dirichlet distribution with parameter vector $\pi = (\delta p'_1, ..., \delta p'_{M!})$, where the parameter δ is a constant that is inversely proportional to the variances of the p_rs, and $(p'_1, ..., p'_{M!})$ is the vector of expected probabilities of the M! rankings, specified by the model of voter behavior. The use of the Dirichlet distribution to generate p_rs for a multinomial process that generates election outcomes leads to the multivariate Pólya (also known as the compound Dirichlet-multinomial) distribution with density function

$$f(N_1, ..., N_{M!}; N, \alpha_1, ..., \alpha_{M!}) = \frac{\Gamma(\delta)}{\Gamma(N+\delta)} \prod_{r=1}^{M!} \frac{\Gamma(N_r + \delta p_r)}{\Gamma(\delta p_r)}, \tag{14}$$

whose first two moments are $E[N_r] = Np_r$, $Var[N_r] = Np_r(1 - p_r)\psi$, and $Cov[N_r, N_s] = -Np_r p_s \psi$, where $\psi = (N + \delta) / (1 + \delta)$.[18] As the variance parameter δ approaches infinity, ψ approaches 1 and the multivariate Pólya distribution in (14) approaches the multinomial distribution in (1). Thus for each model of voter behavior we can estimate the value of δ for which the distribution of *LLR*s or *WSSR*s in simulated elections best matches the distribution in observed elections. The greater the estimated value of δ, the better does the model describe the data.

The standard deviation of the probability p_r as a random variable governed by the Dirichlet process is

$$\begin{aligned} Stdev[p_r] = \sqrt{Var[p_r]} &= \sqrt{\frac{\delta p_r \left(\sum_{i=1}^{M!} \delta p_i - \delta p_r \right)}{\left(\sum_{i=1}^{M!} \delta p_i \right)^2 \left(\sum_{i=1}^{M!} \delta p_i + 1 \right)}} = \sqrt{\frac{p_r \left(\sum_{i=1}^{M!} p_i - p_r \right)}{\left(\sum_{i=1}^{M!} p_i \right)^2 \left(\delta \sum_{i=1}^{M!} p_i + 1 \right)}} \\ &= \sqrt{p_r(1 - p_r)}(\delta + 1)^{-1/2}. \end{aligned} \tag{15}$$

[18] See Mosimann (1962, pp. 67 – 68).

Equation (15) yields an intuitive interpretation of δ. The expression $\sqrt{p_r(1-p_r)}$ on the second line of equation (15) is a semicircle with a radius of 1/2. For the average probability of 1/6, this expression is $\sqrt{5/36}$ = 0.3723. The expression $(\delta+1)^{-1/2}$ can be interpreted as a factor by which this semicircle is compressed. For example, a value of δ = 99 leads to $(\delta+1)^{-1/2}=0.1$, compression by a factor of 10.

The fact that the Dirichlet distribution uses only a single parameter to model the variance of the p_rs leads to variances for the p_rs that are deterministic functions of their means. It is straightforward to accommodate more variability in the variances of the p_rs, at a cost of asymmetry, with a generalized Dirichlet distribution.[19]

4. Empirical Analysis of a Vote-Casting Process

4.1. *The data*

We assemble our ranking data from the "thermometer" scores that are part of the surveys conducted by ANES. These surveys are conducted every two years, and participants are asked to rate politicians on a scale from 0 to 100 (the thermometer). The list of persons includes the president and the vice president, the republican and democratic presidential candidates and vice presidential candidates (in election years), as well as past presidents and presidential candidates who still play significant roles in the political arena. We refer to these persons as "candidates." In the surveys conducted before 1970, a candidate whom the survey respondent did not know received a score of 50 on the participant's answer sheet, while such a candidate was coded as "unknown" in the surveys from 1970 onward. To avoid ambiguities between unknown

[19] See Kotz *et al.* (2000, pp.512-514).

candidates and candidates evaluated at 50, we restrict our analysis to the 18 surveys conducted from 1970 to 2004.

The number of respondents in a survey ranges from 1,212 in 2004 to 2,705 in 1974, and the number of candidates included in the surveys ranges from 3 in 1986 and 1990 to 12 in 1976. We construct all possible combinations of 3 candidates within a year, so we have 1 combination in 1986 and 1990, $12!/(3!\cdot 9!) = 220$ combinations in 1976, and a total of 913 combinations from all 18 surveys. We treat each combination as one election with three candidates. (Recall that we restrict our analysis to three-candidate elections to be able to evaluate our version of the spatial model of voter behavior.) The data from these sets of three candidates are not independent, but we draw no inferences for which an assumption of independence is necessary. It is possible that such three-candidate combinations that are derived from rankings of more than three candidates are qualitatively different from rankings of only three available candidates—for example, because it is often simpler to rank three candidates than a larger number of candidates. However, this problem is likely to be less important for rankings that are derived from thermometer scores, because the ANES survey respondents are asked to evaluate a sequence of candidates on a 101-point scale rather than provide a comparative ranking of all candidates. The thermometer scores are only a small part of the survey, and it is unlikely that the respondents' care in evaluating the candidates changes significantly with the number of candidates they are asked to evaluate (an average of 7.3 candidates per survey). In future research, we plan to investigate whether the statistical properties of three-candidate rankings derived from surveys differ significantly from those of three-candidate rankings in real elections.

For every three-candidate election, we eliminate the survey responses that do not assign scores to all three candidates, yielding from 759 to 2,521 responses for the 913 three-candidate elections. The mean number of responses is 1566.75. For each response, we rank the three candidates according to their thermometer scores. If a response yields a strict ranking of candidates, then we count it as one vote for this ranking. Survey respondents are allowed to assign equal scores to different candidates; about 13 percent of all pairs of responses from a given respondent in our three-candidate elections are two-way ties and about 5 percent of triples of responses are three-way ties. While there are different ways of accommodating ties, we adopt the following intuitive rule: If all candidates are tied, then we count the response as 1/6 vote for each of the 6 possible rankings, and if two candidates are tied, then we count the response as half a vote for each of the two possible strict rankings that break the tie.[20] Thus our adjusted data set consists of the total number of votes for each of the six strict rankings in each of the 913 three-candidate elections. As an aside, we note that 4 among these 913 sets of rankings represent voting cycles.

We first count the frequency with which the three voting methods, Borda, Condorcet, and Estimated Centrality, yield different outcomes in the 913 elections. The three voting methods choose the same winning candidate in 879 elections. This disagreement about the winning candidate in 34 elections (3.7%) suggests that it is worthwhile to inquire as to which voting method is most likely to identify the correct winner. The Borda method and the Condorcet method disagree about the winning candidate in 34 elections, the Borda method and Estimated Centrality disagree in 30 elections, while the Condorcet method and Estimated Centrality disagree in only 14 elections. Thus the Condorcet method and Estimated Centrality agree more often with each other than either of them agrees with the outcome chosen by the Borda method.

[20] We adopt this procedure only to calibrate our models of voter behavior, which requires knowledge of the number of votes for each ranking for the calculation of the mean *LLR* and the mean *WSSR*. The standard methods for accommodating ties in more elegant ways (for example, Black's method—see Black, 1958, pp. 61 – 64) do not allocate shares of tied votes to the respective rankings.

4.2 *Assessing the six models of voter behavior*

The models of equally and unequally likely rankings, the Borda model, the Condorcet model, and the Spatial model of voter behavior impose restrictions on possible vectors of probabilities of using the six strict rankings of three candidates, while the model of equally likely probability vectors implies that every vector of probabilities that sum to 1.0 is feasible. Thus the first five models can be tested by whether observed elections might have been generated by processes that restrict probabilities in the required way, while the model of equally likely probability vectors must be tested by asking whether the data reflect a restriction that is not present in the model.

4.2.1. *Tests of the five models that restrict preference profiles*

We begin by assessing the adequacy of the five models of voter behavior that impose restrictions on the probabilities of the six rankings. In Columns 2 and 3 of Table 3, we report the mean *LLR* and the mean *WSSR*, together with their standard errors, for each of these models. The model of equally likely rankings assumes that the probabilities of all six rankings for all 913 elections are 1/6 and thus uses no degrees of freedom. The model of unequally likely rankings assumes constant probabilities of the six rankings for all 913 elections and thus uses five degrees of freedom. The Borda and the Condorcet models fit a value of α for each of the 913 elections to minimize either the *LLR* or the *WSSR* and thus use 913 degrees of freedom each. The spatial model estimates four parameters (the three angles β_1, β_2, β_3, and the distance x between the mode of the voter ideal points and the triangle's circumcenter) for each of the 913 elections to minimize either the *LLR* or the *WSSR* and thus uses $4 \cdot 913 = 3{,}652$ degrees of freedom. Because of the large differences in their degrees of freedom, we compare the fits of these five non-nested models with the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) in

Table 3. Assessment of six models of voter behavior

		Analysis of observed election data				Analysis of simulated data ("impartial anonymous culture assumption")	
	Degrees of freedom (1)	Mean *LLR* (2)	Mean *WSSR* (3)	AIC (4)	BIC (5)	Mean *LLR* (6)	Mean *WSSR* (7)
Equally likely rankings	0	-196.80	207.28	359,357	359,357	-535.12	581.00
		(4.26)	(4.42)			(0.31)	(0.41)
Unequally likely rankings	5	-31.15	32.45	56,890	56,922	-161.00	160.22
		(0.97)	(0.88)			(0.15)	(0.17)
Borda model	913	-116.79	121.00	215,085	220,951	-367.14	363.84
		(2.72)	(2.78)			(0.23)	(0.24)
Condorcet model	913	-84.99	88.36	157,018	162,885	-297.27	283.20
		(2.00)	(2.25)			(0.20)	(0.20)
Spatial model	3,652	-0.87	0.97	8,893	32,360	-73.15	65.13
		(0.05)	(0.06)			(0.11)	(0.09)

Notes:
1. Standard errors of estimate of the estimated means are shown in parentheses.
2. To facilitate comparisons, we have multiplied the statistics reported in Columns (3) and (7) by 1,000,000.
3. We calculated the AIC and BIC in Columns (4) and (5) using the *LLR*s in Column (2), which share the same denominator. Thus the two measures of fit differ from the conventional measures by an additive constant.
4. To determine the BIC in Column (5), note that there are $5 \cdot 913 - 1 = 4{,}564$ degrees of freedom in the data.

Columns 4 and 5. The *LLR* and the *WSSR* agree with respect to the relative degrees of accuracy of the five models, and we therefore evaluate the AIC and BIC only in terms of the *LLR*s.

The fit of the model of equally likely rankings is far below that of the other four models, which constitutes strong evidence against this model and suggests that, in the typical election, voters report some rankings of candidates significantly more frequently than others. Thus it is meaningful to ask which voting method is most likely to identify the correct winner.

The fit of the Borda model is worse than that of the Condorcet model by about 50 percent, and the fit of the Condorcet model is about two and a half times worse than that of the model of unequally likely rankings. In comparison, the fit of the spatial model is better than that of the other four models by between *one and a half and two orders of magnitude*. Even after accounting for the fact that the spatial model uses far more degrees of freedom than the other models, the results indicate that, by a wide margin, the spatial model describes the observed election data far better than any of the other four models.

4.2.2. *Test of the model of equally likely probability vectors (the impartial anonymous culture assumption)*

We analyze next whether the election data support the impartial anonymous culture assumption that every vector of probabilities p_r from the unit 6-simplex is equally likely. Because this model accommodates every possible outcome perfectly, it requires a different type of evaluation. To test the impartial culture assumption, we draw 1,000,000 samples of probability vectors from a uniform distribution on the unit 6-simplex. For each sample, we draw the number of voters from a gamma distribution with the same mean (1566.75) and standard deviation (380.93) as the number of voters in the election data, and we multiply the six vote shares by the number of voters to obtain the number of votes for each ranking. We then evaluate the 1,000,000 samples by calculating the mean *LLR* and mean *WSSR* for each of the other five models of voter behavior. If the model of equally likely probability vectors is the model of voter behavior that has generated the observed election data, then we would expect the measures of fit of the other five models voter behavior to be similar to those that we obtained from the observed data.

Columns 6 and 7 of Table 1 show the means and the standard errors of estimate of the estimated means for the mean *LLR* and the mean *WSSR* that we calculated from the 1,000,000 samples. For all five models, the estimates of the means of both measures of fit and their respective standard errors of estimate indicate that the ability of any of these models to describe the data generated under the impartial culture assumption is very different from (and also much lower than) its ability to describe the observed data. Thus the hypothesis that the ANES election outcomes were generated by an impartial anonymous culture cannot be sustained.

4.2.3 *Comparisons of distributions that measure the fit of simulated data to observed data*

We next investigate the degree to which our remaining five models of voter behavior are able to simulate election data with the same distribution of *LLR*s as the observed data. Simulating elections with our models of voter behavior requires estimates of the distributions of the unknown model parameters, which we obtain by calibrating the models to the observed data. The model of equally likely rankings does not have any unknown parameters because it simply predicts that each ranking has an equal chance of being chosen. The six shares of our model of unequally likely rankings are the mean shares of the corresponding rankings in our 913 elections.[21] Our estimates of the unknown parameters of the other three models of voter behavior, for each

[21] For example, our estimate of the probability p_1 is the mean share of the rankings with the largest shares in the 913 elections.

election, are the values that maximize *LLR*s of the form of equation (10) for each model. To facilitate replication of our analysis, we report the means and standard deviations of these parameter estimates in Table 4. With respect to the distributions of these parameters, we assume that the unknown scale parameters α in the Borda and the Condorcet models follow gamma distributions whose parameters we calibrate to match the means and variances of the estimates of the αs in Table 4.

Table 4. Estimates of the parameters of three models of voter behavior

Models of voter behavior	Scale parameter α	Angles of the three perpendicular bisectors with the line $\overline{PO}$ β_1	β_2	β_3	Dirichlet variance parameter δ	Distance from O to P x
Borda model	0.3581 (0.1909)					
Condorcet model	0.3603 (0.1871)					
Spatial model		0.5563 (0.3637)	1.5503 (0.4420)	2.5924 (0.3769)	21.5496	0.4447 (0.2290)
Corresponding angles A_i between pairs of bisectors		1.1055 (0.3035)	0.9940 (0.3214)	1.0421 (0.3115)		
Corresponding shares s_i of a semi-circle		0.3519 (0.0966)	0.3164 (0.1023)	0.3317 (0.0992)		
Parameters δ_i implied by means and variances of s_i		23.4326	19.6703	21.5458		

Notes:

(1) Standard deviations are shown in parentheses.

(2) Recall that β_1, β_2, and β_3 specify the three angles formed by the line $\overline{PO}$ and the perpendicular bisectors, which implies that the average angle formed by the two perpendicular bisectors that form the sector that contains O is $A_1 = \beta_1 + \pi - \beta_3$, while the other two angles formed by the perpendicular bisectors are $A_2 = \beta_2 - \beta_1$ and $A_3 = \beta_3 - \beta_2$.

(3) We computed the three estimates δ_i from the means and variances of the three semi-circle shares s_i as $\delta_i = \text{Mean}[s_i] * (1 - \text{Mean}[s_i])/\text{Var}[s_i] - 1$, $i = 1, 2, 3$. The estimate of the Dirichlet variance parameter δ is the mean of the three values δ_i.

For the spatial model, we assume that the three angles between the perpendicular bisectors follow a tri-variate Dirichlet distribution with three parameters (two of the three shares of a semi-circle and the variance parameter δ), and that the distance between O and P follows a Weibull distribution with two parameters. We again calibrate these distributions to match the means and variances of the estimates in Table 4. The fact that the estimates δ_i from the three shares s_i in Table 4 are quite similar suggests that the standard Dirichlet distribution with a single variance parameter δ is likely to provide an adequate description of the variation among the shares.

We undertake the following Monte Carlo experiment for each model of voter behavior: draw a set of parameters from their respective distributions, use them to generate the p_rs, and use the p_rs to draw a set of N_rs from the multinomial distribution. Repeat these steps a large number of times, generating with each repetition a *LLR* from the estimate of the p_rs that produced the simulated N_rs. We then compare the distribution of the *LLR*s that we obtain from the simulations with the corresponding distribution of the *LLR*s that we obtain from the observed elections. For a model of voter behavior that is likely to have generated the observed data, we expect these distributions to be very similar.

To compare these distributions, we order the *LLR*s that we obtained in our analysis of the respective model of voter behavior and choose bounds that distribute the 913 *LLR*s into 11 bins of 83 *LLR*s each. We then allocate the *LLR*s from our evaluation of the simulated data into those bins. The numbers of elections sorted into the bins represent 11 sets of Bernoulli trials, whose distributions can be approximated by the normal distribution for large numbers of trials. Thus the differences between the numbers of observations in the bins from the simulated and the observed elections are asymptotically normally distributed, so that a test statistic equal to the sum of the squared differences between actual and expected numbers of observations divided by the respective variances is asymptotically chi-square distributed with 10 degrees of freedom.[22] Under the hypothesis that the observed election data are generated by the respective model of voter behavior, the right tail area of the chi-square distribution indicates the probability that the test statistic will be greater than or equal to the observed value.

We first test each model of voter behavior with simulated rankings generated by the multinomial distribution in (1). In Column 1 of Table 5, we report the resulting chi-square test statistics, together with the corresponding tail-area probabilities, under the hypothesis that the *LLR*s from the simulated data have the same distribution as the ratios from our analysis of the data from ANES. The large values of the test statistics indicate that none of the four models yields an adequate description of the observed election data. For the model of equally likely rankings, the Borda model, and the Condorcet model, *all* residuals are in the first bin, while 97 percent of the residuals of the model of unequally likely rankings are in the first bin. However, the test statistic of the spatial model is fairly small in comparison to the other four models, which indicates that some version of the spatial model is a more promising candidate for the right model of voter behavior than any of the other models.

[22] This relationship holds only for a random (iid) sample. However, the fact that we demarcate the bins of the observed elections so that the probability of an *LLR* being in any one bin is 1/11 made it straightforward to calculate the correlations between any two bins as a function of the number of voters. We found this correlation to be fairly small—the correlation is ρ = -0.1 if N = 1, ρ = 0.0043 if N = 75, ρ = 0.0071 if N = 150, ρ = 0.0085 if N = 300, and ρ = 0.0093 if N = 600. Thus the correlation increases at a decreasing rate and can be expected to be still below 0.01 if N = 5,000. Because the correlations are very small and our results are fairly strong, we decided that inference based on the chi-square distribution would be sufficiently accurate.

Table 5. Comparison of simulated data with the observed election data

	Multinomial distribution	Pólya distribution		
	Test statistic (p value for 10 df)	δ	Standard deviation of the average p_r	Test statistic (p value for 9 df)
	(1)	(2)	(3)	(4)
Equally likely rankings	11,000,000 (0 %)	18.50	0.0844	23.8100 (0.46 %)
Unequally likely rankings	901,070 (0 %)	68.50	0.0447	20.3062 (1.61 %)
Borda model	11,000,000 (0 %)	25.25	0.0724	24.4329 (0.37 %)
Condorcet model	11,000,000 (0 %)	31.25	0.0656	9.0914 (42.89 %)
Spatial model	175.91 (0 %)	2,439	0.0075	9.6521 (37.93 %)

Notes:
1. The values in parentheses are the chi-square tail-area probabilities under the hypothesis that the distributions of the *LLR*s of the simulated and observed elections are the same.
2. All residuals under the multinomial model for equally likely rankings, the Borda model, and the Condorcet model were concentrated in the first bin.
3. The entries in Column 3 are the Dirichlet standard deviations of p_r = 1/6 that are implied by equation (15).

To create a model that yields *LLR*s with a distribution similar to that of the observed elections, we repeated the analysis under the assumption that the probabilities of the six rankings follow the Pólya distribution in equation (14)—that is, the probabilities are randomized through a Dirichlet process before they become the input of the multinomial distribution. For each model, we estimated the variance parameter of the Dirichlet distribution as the value that minimizes the chi-square test statistic, so the corresponding chi-square test has 11 – 2 = 9 degrees of freedom. We report the estimates of the variance parameters, the standard deviation of the average p_r in the Dirichlet process implied by δ, and the chi square test statistics in Columns 2 – 4 of Table 5. Recall that the variance parameter δ is inversely related to the degree to which the variance of the

Pólya distribution exceeds the variance of the multinomial distribution. Thus the large δ of the spatial model implies that the Dirichlet process contributes only a relatively small amount to the variance of the Pólya distribution, and that the spatial model by itself already describes a large share of the variation in the data. Column 3 indicates that the value of δ in the spatial model implies a standard deviation of the average p_r of about three-fourth of one percent. In contrast, the variance parameters of the four other models are relatively small, indicating that the additional variance in the probabilities that results from the Dirichlet process is fairly large, with standard deviations of the average p_r ranging from 4.4 percent for the model of unequally likely rankings (nearly six times the value for the spatial model) to 8.4 percent for the model of equally likely probabilities (more than 11 times the value for the spatial model). Thus these models of voter behavior explain much smaller fractions of the variation of the *LLR*s of the observed election data, while the Dirichlet error process provides a sizeable share of the necessary variation.

All test statistics in Column 4 are much smaller than those in Column 1, which indicates that the additional Dirichlet process leads to better fits for all models. However, neither the two Pólya models of equal and of unequal probabilities nor the Borda-Pólya model lead to test statistics which would suggest at any conventional level of significance that the simulated data and the observed election data were generated by the same process. In contrast, the tests statistics of both the Condorcet-Pólya model and the spatial-Pólya model suggest at all conventional levels of significance that either model of voter behavior could have generated the observed election data. But because the Dirichlet error process needs to add a much smaller amount of variance to the vote shares of the spatial model than to the vote shares of the Condorcet model, the spatial model of voter behavior, augmented by the specified error process, provides a more satisfying description of the data generating process of the observed ANES election data than the Condorcet model augmented by its error process.

Chamberlin and Featherston (1984) calibrated the Pólya model of unequal probabilities to five presidential elections of the American Psychological Association (APA) between 1976 and 1981. Although they were unable to undertake a meaningful statistical test of their model because they had only five comparison elections, they concluded by eyeballing their simulated data that these "bear a strong structural similarity to the APA elections." Our results for the Pólya model of unequal probabilities suggest that Chamberlin and Featherston's conclusion may seem reasonable on the basis of eyeballing the results (our test statistic of 20.3 *is* reasonably small), but that they may have drawn a different conclusion had they been able to undertake a formal statistical test.

5. Evaluating Voting Methods by Simulation

We use the spatial-Pólya model of voter behavior to estimate the accuracy with which the Borda method, the Condorcet method, and the Estimated Centrality method identify the correct winner. To apply the spatial-Pólya model, we use the distributions that we calibrated to the parameter estimates in Table 4 to first sample the six vote shares from the spatial model, use these shares as input into a Dirichlet process with $\delta = 2{,}439$ to incorporate error into the probabilities, and use these probabilities in a multinomial process with a chosen number of voters—25, 100, 1,000, 10,000, 100,000, and 1,000,000 voters—to obtain the number of votes for each ranking.

For all simulations we assume that the ranking ABC is the correct ranking and that A is therefore the correct winner. The idea of a correct outcome that exists prior to the casting of votes, together with the notion that these votes are random variables, implies that voting can yield an incorrect outcome even when there are just two candidates. Thus when we report the frequency of incorrect outcomes under different voting methods for three candidates, it should be borne in mind that in our model of voting, the possibility of incorrect outcomes does not arise solely because there are three candidates but that this can happen with two candidates as well.

We draw our inference about the accuracy of these voting methods from 1,000,000 simulated elections for each number of voters. We evaluate every simulated election with our three voting methods and count the number of times a voting method does not identify the correct winner. We report the results of this exercise in Table 6. It is worth pointing out that, as a consequence of using the Dirichlet error structure in our generating mechanism, the share of incorrectly identified winners does not converge towards zero as the number of voters increases.

The most striking result is that all three voting methods have very similar error rates for any given number of voters. Thus if the accuracy of the election outcome is the main concern, then, assuming that voters are sincere, the number of voters is more important than the choice of voting method. Nevertheless, the fact that we used 1,000,000 simulations permits us to estimate the error rates of all methods with precision sufficient to identify significant differences among them.[23] The Borda method has a small but statistically significant advantage over the Estimated Centrality method for 25 and 100 voters, while the Estimated Centrality method identifies the correct winner most often if the number of voters is large. Thus with 25 voters, we can expect the Borda method to be ln(0.2794/0.2763) = 1.1 percent less likely to identify the winner incorrectly than the Estimated Centrality method, and 3.4 percent less likely than the Condorcet method. With 1,000,000 voters, the Estimated Centrality method is 3.8 percent less likely than the

[23] We took account of the correlations between the error rates for different voting methods when establishing that for each number of voters, the difference between the lowest and second-lowest error rates is statistically significant.

Condorcet method and 24.4 percent less likely than the Borda method to identify the winner incorrectly.

Table 6. Error rates of voting methods under the (error augmented) spatial model of voter behavior

Number of voters	Borda	Condorcet	Estimated Centrality
25	**0.2763**	0.2858	0.2794
	(0.0004)	(0.0005)	(0.0004)
100	**0.1565**	0.1639	0.1581
	(0.0004)	(0.0004)	(0.0004)
1,000	0.0655	0.0641	**0.0614**
	(0.0002)	(0.0002)	(0.0002)
10,000	0.0457	0.0389	**0.0371**
	(0.0002)	(0.0002)	(0.0002)
100,000	0.0433	0.0353	**0.0338**
	(0.0002)	(0.0002)	(0.0002)
1,000,000	0.0430	0.0350	**0.0337**
	(0.0002)	(0.0002)	(0.0002)

Note: Standard errors are shown in parentheses.

6. Conclusion

The main objective of our paper is to identify the statistical process that governs vote-casting. Our initial analysis suggests that the spatial model of voting, augmented by errors in the probabilities assigned to rankings, is a more suitable description of voter behavior than several other models. Our motivation for identifying this statistical process is to suggest a new direction for evaluating voting methods. Rather than seeking to establish that there are cases in which different voting methods can generate counter-intuitive outcomes, we suggest that it is more promising to analyze how different voting methods perform in actual elections. The goal of such an analysis is to find a voting method that is *on average* more likely than other methods to identify the correct outcome. A group of voters who used such a method in all elections could thus expect to make more correct choices over all elections than if they used other voting methods, even if their method was not necessarily the most accurate in every single election. Such a voting method would arguably be more attractive than the others. Using our model of voter behavior, we show that the Borda method (for small numbers of voters) and the Estimated Centrality method (for large numbers of voters) are somewhat more likely to identify the correct winner in actual elections than the Condorcet method.

We emphasize that our current analysis is far from being conclusive; it constitutes only the beginning of a larger research program with at least four different lines of inquiry. First, we draw our conclusions in this paper on the basis of a single data set that is derived from surveys. We need to test whether our results will continue to hold for data derived from real elections, and whether it is even possible to describe all elections with a single model of voter behavior. Chamberlin and Featherston (1985) provide some evidence that ANES survey data have similar properties as their data from the presidential elections of the APA, but that the statistical properties of elections that take place in the same environment tend to be more similar than those of elections from different environments. Still, the fact that the spatial model fits the ANES data so much better than the four competitors that we examine suggests that the spatial model may very well continue to be a reasonable description of voter behavior in other elections.

Second, it is probably possible to improve upon the spatial model that we use. The fact that we did not obtain a satisfying fit of our spatial model until we introduced some—albeit small—additional variation through a Dirichlet process indicates that our spatial model does not yet describe voter behavior perfectly. We have only examined a very simple version of the spatial model, and a more general spatial model might fit the observed data even better.

Third, we can use our statistical framework to investigate the accuracy of different voting methods. There are many popular voting methods that are either widely used or frequently suggested as alternatives to current methods, and it would be valuable to know the frequency with which each method is able to identify the correct winner. The fact that our preliminary analysis has identified the Borda method and the Estimated Centrality method as the most accurate for different sizes of the electorate does not imply that these methods are necessarily more accurate than others.

Finally, our analysis has implications for all inquiries into the frequency of certain voting events, be that the probability that strategic voting will alter the outcome of an election, the existence of dominant candidates, or the likelihood of different voting paradoxes. Rather than assuming that either all rankings or all vectors of probabilities are equally likely (the *impartial culture* and *impartial anonymous culture* assumptions), our framework makes it possible to incorporate more realistic statistical models of voter behavior into such analyses and thereby to improve the accuracy of their predictions. In this context it is worth reiterating that our results regarding the accuracy of voting methods are based on the assumption that voters do not vote strategically. Any recommendation about which voting method one ought to use in actual elections needs to take the method's resistance to strategizing into account. Fortunately, our framework quite naturally lends itself to such an inquiry.

References

Adams, James. 1997. "Condorcet Efficiency and the Behavioral Model of the Vote." *The Journal of Politics* **59**:4, 1252-1263.

Adams, James. 1999. "An Assessment of Voting Systems under the Proximity and Directional Models of the Vote." *Public Choice* **98**:131–151.

Berg, Sven. 1985. "Paradox of Voting under an Urn Model: The Effect of Homogeneity." Public Choice 47:2, 377 – 387.

Black, Duncan. 1958. *The Theory of Committees and Elections*. Cambridge: Cambridge University Press.

Brams Stephen J, and Peter Fishburn. 2001. "A Nail-biting Election." *Social Choice and Welfare* **18**: 409–414.

Campbell, Colin and Gordon Tullock. 1965. "A Measure of the Importance of Cyclical Majorities." *Economic Journal* **75**:300, 853-857.

Cervone, Davide; William Gehrlein; William Zwicker. 2005. "Which Scoring Rule Maximizes Condorcet Efficiency Under Iac." *Theory and Decision* **58**:2, 145-185.

Chamberlin, John R. and Michael D. Cohen. 1978. "Toward Applicable Social Choice Theory: A Comparison of Social Choice Functions under Spatial Model Assumptions." *American Political Science Review* **72**:4, 341-1356.

Chamberlin, John R., Jerry L. Cohen, Clyde D. Coombs. 1984. "Social Choice Observed: Five Presidential Elections of the American Psychological Association." *The Journal of Politics* **46**:2, 479-502.

Chamberlin, John R. and Fran Featherston. 1986. "Selecting a Voting System." *The Journal of Politics* **48**:2, 347-369.

Condorcet, Marie Jean Antoine Nicolas de Caritat, Marquis de. 1785. Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix (Paris).

Conitzer, Vincent and Tuomas Sandholm. 2005. "Common Voting Rules as Maximum Likelihood Estimators." In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI-05)*, pp. 145-152, Edinburgh, Scotland, UK.

DiDonato, A.R. and R.K. Hageman. 1980. "Computation of the Integral of the Bivariate Normal Distribution Over Arbitrary Polygons." Naval Surface Weapons Center, Government Accession Number ADA102466.

Drissi-Bakhkhat, Mohamed and Michael Truchon. 2004. "Maximum Likelihood Approach to Vote Aggregation with Variable Probabilities" *Social Choice and Welfare* **23**: 161-185.

Enelow, James M. and Melvin J. Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. Cambridge University Press.

Enelow, James M. and Melvin J. Hinich. 1990. *Advances in the Spatial Theory of Voting*. Cambridge University Press.

Gehrlein, William V. and Fishburn. 1976. "Condorcet's Paradox and Anonymous Preference Profiles." *Public Choice* **26**:1-18.

Gehrlein, William V. 2002. "Condorcet's Paradox and the Likelihood of Its Occurrence: Different Perspectives on Balanced Preferences." *Theory and Decision* **52**:2, 171-199.

Good, I. Jack and T. Nicolaus Tideman. 1976. "From Individual to Collective Ordering through Multidimensional Attribute Space." *Proceedings of the Royal Society of London* (Series A) **347**: 371-385.

Hendry, David. 1980. "Econometrics: Alchemy or Science?" *Economica* **47**: 387-406.

Kemeny, John. 1959. "Mathematics without Numbers." *Daedalus* **88**: 571-91.

Kendall, Maurice and Jean D. Gibbons. 1990. Rank Correlation Methods. New York: Oxford University Press.

Kuga, Kiyoshi and Hiroaki Nagatani. 1975. "Voter Antagonism and the Paradox of Voting." *Econometrica* **42**: 1045-1067.

Kotz, Samuel, N. Balakrishnan, Norman L. Johnson. 2000. *Continuous Multivariate Distributions, Volume 1, Models and Applications*, Wiley-Interscience (2nd Edition).

Merrill, Samuel. 1984. "A Comparison of Efficiency of Multicandidate Electoral Systems." *American Journal of Political Science* **28**: 23-48.

Mosimann, James E. 1962. "On the Compound Multinomial Distribution, the Multivariate Beta Distribution, and Correlations Among Proportions." *Biometrika* **49:** 65-82.

Regenwetter, Michel and Bernard Grofman. 1998a. "Approval Voting, Borda Winners, and Condorcet Winners: Evidence from Seven Elections." *Managment Science* **44**:4, 520-533.

Regenwetter, Michel and Bernard Grofman. 1998b. "Approval Voting, Borda Winners, and Condorcet Winners: Evidence from Seven Elections." *Social Choice and Welfare* **15**:423-443.

Regenwetter, Michel, Bernard Grofman, and Anthony A.J. Marley. 2002. "On the Model Dependence of Majority Preference Relations Reconstructed from Ballot or Survey Data." *Mathematical Social Sciences* **43**: 451 – 466.

Regenwetter, Michel and Ilia Tsetlin. 2004. "Approval Voting and Positional Voting Methods: Inference, Relationship, Examples." *Social Choice and Welfare* **22**: 539-566.

Risse, Mathias. 2001. "Arrow's Theorem, Indeterminacy, and Multiplicity Reconsidered." *Ethics* **111**: 706-734.

Risse, Mathias. 2005. "Why the Count de Borda Cannot Beat the Marquis de Condorcet." *Social Choice and Welfare* **26**: 95-113.

Saari, Donald G. 1990. "Susceptibility to Manipulation." *Public Choice* **64**: 21-41.

Saari, Donald G. 1999. "Explaining All Three-Alternative Voting Outcomes." *Journal of Economic Theory* **87**: 313-355.

Saari, Donald G. 2001. "Analyzing a Nail-biting Election." *Social Choice and Welfare* **18**: 415–430.

Saari, Donald G. 2003. "Capturing the 'Will of the People.'" *Ethics* **113**: 333-349.

Saari, Donald G. 2006. "Which is Better: The Condorcet or Borda Winner?" *Social Choice and Welfare* **26**: 107-129.

Spanos, Aris. 1986. *Statistical Foundations of Econometric Modeling*. Cambridge.

Tideman, T. Nicolaus. 2006. *Collective Decisions and Voting*. Burlington, VT: Ashgate.

Truchon, Michel. 2006. "Borda and the Maximum Likelihood Approach to Vote Aggregation." *Mimeo*.

Tsetlin, Ilia and Michel Regenwetter. 2003. "On the Probabilities of Correct or Incorrect Majority Preference Relations." *Social Choice and Welfare* **20**:283 – 306.

Tsetlin, Ilia, Michel Regenwetter, and Bernard Grofman. 2003. "The Impartial Culture Maximizes the Probability of Majority Cycles." *Social Choice and Welfare* **21**:3, 387-398.

Young, H. Peyton. 1986. "Optimal Ranking and Choice from Pairwise Comparisons." In Grofman B. and G. Owen (eds.) Information Pooling and Group Decision Making. Greenwich, Conn: JAI Press, pp. 113-122.

Young, H. Peyton. 1988. "Condorcet's Theory of Voting." *The American Political Science Review* **82**(4): 1231-1244.

Young, H. Peyton. 1995. "Optimal Voting Rules." *Journal of Economic Perspectives* **9**:1, 51-64.