International Institute for
Applied Systems Analysis
IIASA www.iiasa.ac.at

The Source of Some Paradoxes from Social Choice and Probability

Saari, D.G.

IIASA Collaborative Paper
August 1986

THE SOURCE OF SOME PARADOXES FROM SOCIAL CHOICE AND PROBABILITY

Donald G. Saari

August 1986
CP-86-25

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

FOREWORD

This paper offers a powerful, simple method for understanding many "paradoxes" in social choice and probability theory. The approach is a geometrical one; the underlying principle emerges from a wide variety of examples ranging from elections and agenda manipulation to gambling and conditional probabilities.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

THE SOURCE OF SOME PARADOXES FROM SOCIAL CHOICE AND PROBABILITY

Donald G. Saari

Department of Mathematics, Northwestern University, Evanston, Illinois

1 INTRODUCTION

The social choice literature and the probability literature are filled with descriptions of "paradoxes". As we show here, many of them can be explained and extended by using the same, simple, geometric argument. Our extensions include several new results about the intransitivities of election results over subsets of alternatives, the cycles of agenda manipulation, gambling, and Simpson's paradox from conditional probability. Furthermore, we prove that these paradoxes must accompany the modeling in a robust fashion. Our approach appears to be new, it is elementary (based on the open mapping principal), and it uncovers new examples. Indeed, one point that emerges is the ease with which paradoxes (i.e., apparent contradictions in a relationship) can arise.

Our argument extends beyond social choice and probability, but we emphasize these two areas because of their familiarity and their importance as standard modeling tools for economics and decision analysis. Examples from probability are discussed in Sections 2 and 3; examples from social choice are discussed in Section 3. To simplify the exposition, we use discrete random variables. However, all this work easily generalizes to more general models.

The flavor of our results is indicated by the following two prototype examples. In the following sections, we show how they are related and how they can be extended.

1.1 Conditional Probability and Simpson's Paradox

Suppose a certain drug is tested in Chicago (C) ad in Los Angeles (C'). A test group (T) receives the new drug, and a control group (T') received the standard treatment. Some people are returned to health (H), while others are not (H'). Suppose that in both communities the new drug is judged to be successful because it cures the sick with a higher ratio than the standard treatment

$$\{P(H{:}CT) > P(H{:}CT'),\ P(H{:}C'T) > P(H{:}C'T')\} .$$

Is it possible for the aggregated test results to have the reverse conclusion $P(H{:}T) < P(H{:}T')$? It is, and this is known as Simpson's paradox. An explanation (which differs from that given here) and how it relates to the "sure thing" principle is given by Blyth (1972a). Examples with real data are given by Wagner (1982).

This inconsistency phenomenon turns out to be a characteristic of models based on conditional probability or the combination of random variables. When additional conditions are introduced, almost any imaginable extension can occur. For instance, suppose that the tests are conducted using facilities provided by universities (U) and private laboratories (U'). There exist examples whereby the new drug is unsuccessful at each university and at each laboratory as well as in each community, but it is successful in the aggregate, and there exist examples where the conclusions oscillate with the level: the new drug is successful at each of the facilities, it is unsuccessful in each community, and it is successful in the aggregate, etc.

The appropriate ratio of success to failure in each of these examples can be made to exceed any predetermined constant. This means that there exist examples of data where at each facility the probability of regaining health by use of the drug is at least twice that obtained by the standard treatment; in each community, the standard treatment is at least three times better than the drug; and at the total aggregated level, the drug is at least four times better than the standard treatment!

1.2 Voting and Ranking Methods

The aggregation of preference is a central issue in the social sciences. A simple system is voting. Here, several paradoxes occur when the voters rank a set of three alternatives $\{a, b, c\}$ by using the standard plurality voting system. Suppose there are nine voters where four of them giv the ranking $c > a > b$, three give the ranking $b > a > c$, and two give the ranking $a > b > c$. The group's ranking is $c > b > a$ with a tally of 4 : 3 : 2. This ranking is inconsistent with the fact that a majority of the voters (five of them) prefer the bottom ranked alternative a to the top ranked alternatives c.

It might be suspected that an election ranking of N alternatives must have some relationship to how the group ranks at least one of the pairs of alternatives. This need not be the case. For each of the $N(N-1)/2$ pairs of alternatives designate, in an *arbitrary* fashion, one of the alternatives. We show that there exist examples of voters' rankings of alternatives so that (1) the plurality election result is $a_1 > a_2 > \ldots > a_N$, and (2) for each pair of the alternatives, a majority of the *same* voters prefer the designated alternative.

2. THE GENERAL RESULT

The simple geometric property where open sets are mapped to open sets is the unifying explanation for all th paradoxes described in this paper. The following standard statement (see Warner, 1970) suffices for what follows.

Proposition 1

Let F be a smooth mapping from an m-dimensional manifold M to an n-dimensional manifold N where $m > n$. Let c be an interior point of N. Assume that p in $F^{-1}(c)$ is an interior point of M. If the Jacobian of F at p has maximal rank, then there is an open neighborhood of p that is mapped onto an open neighborhood of c.

The proof of the following theorem illustrates why the above property is the source of the paradoxes.

Theorem 1

Consider the example in Section 1.1 where a drug is compared with a standard treatment. Let A be the variable representing the seven sets $C + C', C, C', CU, CU', C'U$, and $C'U'$. For each choice of A, designate which term from the pair $(P(H:TA), P(H:T'A))$ is to have the larger value. Choose a constant d_A greater than unity and express each pair as a ratio that is bounded below by d_A. There exist finite examples of data that simultaneously satisfy all the seven specified inequalities.

To prove the theorem, it suffices to show that, for any choice of signs for the seven quantities $P(H{:}TA) - P(H{:}T'A)$, there exist sample points that realize them. View these quantities as defining the seven components of a mapping F into R^7. The choice of the signs identifies an orthant B of R^7.

The origin O of R^7 is a boundary point for each of the orthants. This "comparison point" is used in the following way. First, a point p in $F^{-1}(0)$ is found so that (i) it is an interior point of the domain, and (ii) the Jacobian of F at p has maximal rank. According to Proposition 1, F maps an open neighborhood of p onto an open neighborhood of O. This open image set meets each of the orthants; in particular it meets orthant B. Therefore, there are sample points that satisfy all seven conditions simultaneously. The technical part of the proof is to define the domain so that F can be represented by a smooth mapping.

Proof

There are eight sets determined by the intersections of the sets T, C, U and their compliments. They are:

$$S_1 = TCU \quad S_2 = TCU' \quad S_3 = TC'U \quad S_4 = TC'U'$$

$$S_5 = T'CU \quad S_6 = T'CU' \quad S_7 = T'C'U \quad S_8 = T'C'U'$$

Treat each of these sets as a disjoint space. Let X_j designate the characteristic function of H in S_j. Define $Y_j = X_{2j} + X_{2j-1}, j = 1,\dots,4$, and $Z_j = Y_{2j} + Y_{2j-1}, j = 1,2$. The random variable Y_j, which is the characteristic function of H over $S_{2j} + S_{2j-1}$, represents the results at the community level, while $Z_j = Y_{2j} + Y_{2j-1}$ represents the final aggregated results.

If x_j denotes the value of $P(X_j = 1)$, then x_j is in the unit interval $I, j = 1,\dots,8$. Let d_j designate $P(S_j)$ in the space $\cup S_k$. The d_j variables describe a simplex in R^8 which is denoted by Si (8) and defined by

$$\sum_{j=1}^{8} d_j = 1 \ .$$

These 16 variables are in the 15-dimensional space $M = I^8 \times Si\,(8)$.

By use of the standard relationship, for any set E,

$$P(B) = P(BE) + P(BE') = P(E)P(B{:}E) + P(E')P(B{:}E') \tag{2.1}$$

it follows that the probabilities $y_j = P(Y_j = 1)$ and $z_j = P(Z_j = 1)$ are the rational functions

$$y_j = \frac{x_{sj-1}d_{2j-1} + x_{2j}d_{2j}}{d_{2j-1} + d_{2j}}$$

$$z_1 = \sum_{j=1}^{4} x_j d_j / \sum_{j=1}^{4} d_j$$

and

$$z_2 = \sum_{j=5}^{8} x_j d_j / \sum_{j=5}^{8} d_j$$

Comparison map: Let $F:M \to R^7$ be

$$F = \sum_{j=1}^{2} \{x_j - x_{j+4}\}e_j + \sum_{j=1}^{2} \{y_j - y_{j+2}\}e_{j+4} + \{z_1 - z_2\}e_7 \qquad (2.2)$$

where $\mathbf{e}_j$ is the unit vector in R^7 with unity in its j^{th} component. The components of F represent $P(H:TA) - P(H:TA')$ as A ranges through its seven values.

Open mapping: Clearly, F is a smooth mapping. That the Jacobian of F has maximal rank at some preimage of O is a direct computation. Indeed, this rank condition holds everywhere except on a certain lower-dimensional subset of M. These points of lower rank are where either the y values or the z values are uniquely determined because the corresponding pairs of x or y are equal.

The signs chosen for the seven quantities determine an orthant of R^7, denoted by B. By construction, all the sample points with this behavior are in $U = F^{-1}(B)$. By the continuity of F,U is an open set; we must show that it is nonempty. The Jacobian of F has maximal rank at some interior point of M in $F^{-1}(0)$, so F maps an open set from M onto an open set of O. This open set meets B. Consequently $F^{-1}(B)$ is nonempty.

Next we show that U contains points that can be identified with a finite data set. Any rational point will suffice. A multiple of the common denominator of $\mathbf{d}$ is the total number of subjects. The same multiple of the numerator of $\mathbf{d}$ corresponds to the cardinality of S_j, and it serves as a multiple of the denominator of x_j. Because the rational points are dense, there is an infinite set of rational points in $F^{-1}(B)$. Each point can be identified with an infinite number of different finite data sets.

The set $F^{-1}(O)$: It remains that the inequalities can be bounded below by the designated constants. Once the values of d_A are specified, the inequalities are of the type $x_j > d_A X_{j+4}$ with a similar relationship y and z. Let $q = (q_1, \ldots, q_7)$ be a point in B. The set $F^{-1}(q)$ is given by equations of the form $x_j - x_{j+4} = |q_j|$ with similar equations for y and z. These equations define lower-dimensional hyperplanes in the domain, so there are points in U satisfying inequalities of the form

$$\frac{x_j}{x_{j+4}} > 1 + \frac{|q_j|}{x_{j+4}} \tag{2.3}$$

The assertion follows if there is a point in U such that the right-hand sides of these inequalities are bounded below by d_A, with similar statements for y and z. These inequalities are satisfied if the values of the denominators on the right-hand sides can be chosen to be arbitrarily small. This involves a direct computation that is easily done because $F^{-1}(O)$ contains the intersection of $O \times Si(8)$ and the boundary of U. The points are chosen arbitrarily close to this set.

Comments:

(1) The basic idea of this proof extends to all the paradoxes discussed here. Individual comparisons are one dimensional. When several comparisons are made, they *must* be viewed as defining a comparison mapping F with a higher-dimensional range space. A higher-dimensional space admits symmetries and cycles, so it should be expected that these cycles are manifested as paradoxes by the comparisons. To prove that all the symmetries are admitted, locate a "comparison point" on the boundary of each of the comparison regions. Next, show that the image of F includes an open set about the comparison point. The intersection of this open set with each comparison region is a nonempty open set. Because F is continuous, this means that there is a nonempty set of points in the domain with the desired properties. To complete the proof, impose conditions so that in each of these sets in the domain there exist points that are identified with sample points from the model. This simple idea is the essence of our explanation for all the paradoxes in this paper.

(2) The number of possible, paradoxial relationships is determined by the dimension of the domain for a comparison mapping. If this dimension exceeds that of the range, then the comparison mapping is not "complete"; there exist additional relationships that may define more complex

paradoxes. As a corollary, the above illustration and extension of Simpson's paradox is *not* the "best possible" result. The domain of F is 15 dimensional while the range is only seven dimensional; eight more comparisons using these variables can be added. (They may involve different levels of aggregation, waiting times, etc.)

(3) Other conclusions are derived from the properties of F^{-1}.

(i) It is natural to determine the limits of a paradox. (This is illustrated in Theorem 1 with the assertion that d_A is not bounded above.) Often, as for this model, these limits are determined by the properties of the points near the intersection of $F^{-1}(O)$ and the boundary of the domain.

(ii) In order for a paradox (described by B) to occur, we may need a certain number of data points. The minimal size is given by the smallest "lowest common denominator"of the admissible points in $F^{-1}(B) = U$.

(iii) The probability that a paradox (described by B) occurs is given by the measure of a probability distribution over the open set $F^{-1}(B)$.

(4) Other conclusions are derived from the structure of the image of F. For instance, the image contains an open set about the origin, so it meets any sector defined by a specified ratio of the outcomes; e.g.,

$$[(P(Y_1 = 1) - P(Y_3 =)]^3 > 42[P(X_1 = 1) - P(X_4 = 1)] > 0$$

The above shows that there are sample points that satisfy these conditions.

(5) To avoid the above behavior, the Jacobian of F cannot be of maximal rank. This singularity constraint becomes a ***necessary*** condition to avoid a paradox. Often, as for the above model, these lower-dimensional singularity conditions correspond to familiar constraints such as the "independence of random variables".

This approach can be used as long as the components of a comparison mapping are smooth functions. These components could be functional combinations of probabilities, expected values, the various moments, waiting times, loss functions, decision rules, correlation indices, scattering indices, covariance, etc. If the open mapping condition holds at a comparison point, then all possible comparisons are realized. In this way it is easy to show that there exist examples illustrating, or instance, that the expected value may satisfy $E(X) > E(Y)$, yet $E(f(X)) < E(f(Y))$ for some monotonically increasing function f, and that certain decision rules may be inconsistent with other measures. (Indeed, a discrete version of this can be used to explain the Arrow social choice paradox.) Theorem 2 is the formal statement that covers all these situations.

Before stating Theorem 2, we formally define the structural relationship between a comparison point and a comparison region.

Definition

Let a topological space N be partitioned. A ***comparison point*** is a boundary point for each partition set. For a given comparison point p, a ***comparison region*** is a partition set such that the closure of its interior contains p.

Definition

Let $F{:}M \to N$ be a comparison mapping for a given model. A point in M is an ***admissible point*** if it can be identified with a sample for the model.

Theorem 2

Let $F{:}M \to N$ be a smooth comparison mapping where the dimension of M is bounded below by the dimension of N. Assume that the admissible points form a dense set in M. Let c be a comparison point in N. If p in $F^{-1}(c)$ is an interior point of M such that the Jacobian of F at p has maximal rank, then the behavior characterized by any comparison region of N is admitted.

Example

Consider the following dice game. Each of the players rolls his own weighted die. (Each die is marked in the standard fashion.) On each roll, the winner is the player that rolled the larger face value. For each choice of $k = 1,\ldots,4$, the losing player pays the winning player the difference between the face values raised to the k^{th} power. For each of the four choices of k, arbitrarily select a die to have the larger expected payoff, and then arbitrarily select a die to have the higher probability of winning a roll. It is a direct consequence of Theorem 2 that the dice can be weighted in such a fashion that all five selected conditions are satisfied simultaneously. This illustrates the possible incompatibil:.y among reward functions and the distributions.

As a special case ($K_1 = 1$), the following demonstrates that the more probable of two events may have the longer waiting time.

Corollary 2.1

Let each of the two urns U_1 and U_2 contain red and black balls. Each urn is randomly sampled without replacement. For a positive integer k, let $w_j(k)$ be the probability that it takes at least k tries before a red ball is selected from urn U_j, $j = 1,2$. Let $k_1 \neq k_2$. For each of the two pairs $(w_1(k_s), w_2(k_s)), s = 1,2$, choose the value that is to be the larger. There exist examples of data so that both conditions are satisfied simultaneously.

Outline of the proof

We proof the special case where $K_1 = 1$ and $K_2 = 2$. The general case follows in much the same manner.

The domain of the comparison mapping F is $I \times I \times R_+$ where R_+ is the half-line of positive numbers. A point in the domain is denoted by (x, y, z). Let

$$F = [x - y, x + \frac{(1-x)xzn}{zn-1} - \{y + \frac{(1-y)yn}{n-1}\}]$$

be a mapping into R^2. At the rational points in the domain, F can be identified with the mapping $(w_1(1) - w_2(1), w_1(2) - w_2(2))$. This identification follows by assuming that there are xzn red and $zn(1-x)$ black balls in U_1 and yn red and $n(1-y)$ black balls in U_2, and by choosing an appropriate value for the parameter n.

The comparison point is $0 = (0,0)$. Any domain point p of the form $(x, x, 1)$ is in $F^{-1}(0)$. The gradient of the first component of F is $(1, -1, 0)$; the gradient of the second component evaluated at p is

$$[1 + \frac{(1-2x)zn}{zn-1}, -\{1 + \frac{(1-2x)zn}{zn-1}\}, \frac{(x-1)x}{(zn-1)^2}]$$

where $z = 1$. Clearly, these two vectors are linearly independent. This completes the proof.

The conclusion of this corollary holds even if one of these pairs is replaced with the pair of expected waiting times. However, the conclusion does not hold if the sampling is with replacement, or if the number of balls is the same for each urn. For each of these models, the z term does not appear in the definition of F. As a result, the third component of the gradient is zero, and the second is the negative of the first. Therefore, the Jacobian of the comparison mapping is singular. This illustrates comment (5).

A more interesting paradox is obtained by combining the model in Theorem 1 with the one given above. Here, several pairs of urns with red and black balls are used. As the contents of the urns are combined in a specified way, the urn with the higher probability of selecting a red ball may change with the level of aggregation. Furthermore, the waiting time to select a red ball may vary. However, as demonstrated above, to obtain these examples, often we need the extra degrees of freedom offered by varying the number of balls per urn. Also, the number of independent comparisons is bounded by the dmension of the domain.

Several interesting paradoxes from population dynamics involve only a small number of comparisons, so it is to use Theorem 2 to explain and extend them (see, for example, the paper by Vaupel and Yashin, 1985). However, often these examples, as given by Vaupel and Yashin, are based on continuous random variables. To use Theorem 2, the continuous variables are approximated by discrete valued random variables. Alternatively, Proposition 1 can be extended, in the obvious fashion, to permit M to be a function space. In this way, the examples of Vaupal and Yashin can be treated directly.

Another source of paradoxes subsumed by Theorem 2 is Blyth's paper (1972 b). One of his paradoxes with random variables X and Y has $P(X > Y)$ as close to unity as desired, even though $P(X < \alpha) < P(Y < \alpha)$ for all choices of α. This, of course, is an example of the boundary behavior of the comparison mapping. Both Blyth (1972 b) and Vaupel and Yashin (1985) describe the paradoxes in terms of examples. The above treatment explains and unites them, it shows that they can be extended in several ways, and it proves that the paradoxes are "robust" in that they are satisfied by open sets of examples.

Theorem 1 and its generalization to a set of N characteristic functions are corollaries of Theorem 2. The only surprising feature of the generalization is that the dimension of the domain for a comparison mapping can be very large. To see this, let the first $N-1$ characteristic functions define 2^{N-1} sets. The last random variable is treated as a characteristic function on each set. Thus, the domain of a comparison mapping has the dimension $2^{N-1} + (2^{N-1} - 1) = 2^N - 1$. According to comment (2) this means that up to $2^N - 1$ functional relationships can be defined from these random variables with possible concomitant unexpected behavior.

We conclude this section with a partial converse for Theorem 2. It asserts that if a certain set of examples can be found, then examples of all types exist. Such a result is of value because when the dimension of the range space is sufficiently large it may be difficult to verify the rank condition. However, the specified set of examples might be identified by a computer search. For simplicity, we restrict attention to linear comparison maps.

Corollary 2.2

Suppose that F is a linear comparison mapping from a linear space to a range space R^k. Assume that the 2^k orthants of R^k are comparison regions. If there exist $2^{(k-1)}+1$ examples, each in a different comparison region, then F has maximal rank and all possible comparisons are admitted.

Proof

The image of a linear space under a linear mapping is a linear space. If this image space has dimension k, then the conclusion follows. By assumption, k image points can be found that do not lie in the same $(k-1)$-dimensional subspace. This completes the proof.

Extensions are obvious. For instance, the proof requires only k examples that are not in the same $(k-1)$-dimensional plane. For other choices of comparison regions, the emphasis is placed on the geometry defined by the image points with respect to the properties of the image set.

3 RANKING PARADOXES

A richer assortment of paradoxical behavior emerges from multivalued random variables. (This is because the dimension of the domain for a comparison mapping increases with the number of values admitted by a random variable.) We illustrate this with several new results about ranking and voting procedures.

Our main results concern weighted or positional voting. This is defined in the following way. To rank the N alternatives, $a_1,\dots,a_N$, choose N scalar weights $(w_1,\dots,w_N)$ where $w_j \geq w_k$ if and only if $j < k$ and where $w_1 > w_N \geq 0$. Each voter lists his ranking of the N alternatives on a ballot. To tally a ballot, w_j points are assigned to the j^{th} ranked alternatives, $j = 1,\dots,N$. In the obvious way, the group's ranking of the alternatives is determined by the sum of the assigned weights.

The weights define a voting vector $W_N = (w_1,\ldots,w_N)$ in R^N. For plurality voting, the voting vector is $(1,0,\ldots,0)$. Another well-known voting system, called the Borda count, is defined by the voting vector $B_N = (N, N-1,\ldots,1)$. (If a voting vector is a linear combination of B_N and $E_N = (1,\ldots,1)$, then we call the system a Borda system. This is because the election result for any Borda system always agrees with the result when B_N is used to tally the ballots (see Saari (1982.))

This tallying process can be identified with the expected value of multivalued random variables. For N alternatives, there are $N!$ different ways to rank the N alternatives. Since the sum assigned to any alternative is a linear relationship, the group's ranking is not altered should each sum be divided by the total number of voters. This means that the number of voters is replaced with the fraction of the voters with each ranking. In this way the domain for this problem can be identified with (the rational points in) the simplex $Si(N!)$. The simplex is in the positive orthant of an $N!$-dimensional space. If A_j is the random variable assigned to alternative a_j, then $P(A_j = w_k)$ denotes the fraction of the voters who rank the j^{th} alternative in the K^{th} place. Let A be the vector valued random variable $(A_1,\ldots,A_N)$. A point in $Si(N!)$ can be viewed as being a probability distribution, and so the tally of the ballots can be identified with the expected value $E(A)$.

The following definitions are used in what follows.

Definition

A voters' *profile* is a listing of each voter's ranking of the N alternatives.

Definition

The voting vector W_N defines a *reverse neutral* system if $W_N + (w_N,\ldots,w_1) = CE_N = (C,\ldots,C)$ for some scalar C.

A Borda system is always reverse neutral. An easy algebraic argument demonstrates that the space of reverse neutral systems is a hyperplane of R^N with dimension $1 + [N/2]$ where [] denotes the "greatest integer function". A basis for this hyperplane can be computed directly. For $N = 3$ only the Borda systems are reverse neutral. For $N = 4$, a basis for the hyperplane is given by E_4, and (2,1,1,0). For $N = 5$, a basis is E_5, B_5 and (2,1,1,1,0); etc.

Although weighted voting systems are an important class of voting methods, the interpretation of election results is problematic. This is illustrated by the following theorem which includes Example 2 as a special case.

Theorem 3

Let $N \geq 3$. Let Al_k be the subset $\{a_l,\ldots,a_k\}$. Let W_k be the voting vector used to rank Al_k. Assume that W^k is not reverse neutral, $k = 3,\ldots,N$. For each of $N(N -)/2$ pairs of alternatives, arbitrarily designate one of the alternatives. Arbitrarily choose a ranking RK_k for the set $Al_k, k = 3,\ldots,N$. There exist profiles of voters such that (i) the election result for Al_k is $RK_k, k = 3,\ldots,N$, and (ii) for each pair of alternatives, a majority of these same voters prefer the designated alternative.

The proof of this theorem and an extension are given in Section 4.

A simple consequence of this theorem is that, if a Borda system is not used, then there are profiles of voters where most of the voters prefer a_1 to a_2, most prefer a_2 to a_3, most prefer a_1 to a_3, yet the election result is $a_3 > a_2 > a_1$. The implied ranking obtained by majority vote over the pairs of alternatives is the reversal of the election result! (In the example in Section 1.2, such a profile is given for plurality voting.) A more striking example is that for $N = 5$ there is a profile of voters such that majority votes determine the rankings $a_j > a_{j+1}$ for $j = 1,\ldots,4$, $a_5 > a_1$ {these five alternatives from a cycle} $a_4 > a_j$ for $j = 1,2, a_1 > a_3$, and $a_j > a_5$ for $j = 2,3$, and the plurality election results of $Al_j, j = 3,4,5$, are $a_1 > a_3 > a_2, a_2 > a_3 > a_1 > a_4$, and $a_3 > a_1 > a_5 > a_2 > a_4$ respectively. Other examples are limited only by the imagination of the designer.

By use of different techniques, Fishburn (1981) proved Theorem 3 for the special case $N = 3$. (More accurately, Fishburn gave a proof only for the first example above. However, it is possible that his approach extends to include our general statement for $N = 3$.) For $N > 3$, the first conclusion without the second is a special case of a result given by Saari (1984).

The second part of the theorem is of independent interest. Essentially, it asserts that if the pairs of alternatives are ranked by majority voting, then any type of cycle, subcycle, etc., can occur. To highlight this result, we restate it.

Definition

Let N => 3. For $1 \leq k < j \leq N$, let R_{kj} be the set $\{a_k > a_j, a_j > a_k\}$. Let the space of binary rankings BR be the cartesian product of the $N(N-1)/2$ sets R_{kj}.

An element of BR is a sequence that imposes an ordering for each of the $N(N-1)/2$ pairs of alternatives. These binary rankings need not be transitive, nor need they satisfy any other consistency requirement.

Corollary 3.1

Let q be an element BR. There exist examples of voters' profiles such that, for each pair of alternatives, a majority of the same voters have the ranking specified by q.

The remainder of this section is devoted to extracting some of the consequences of Theorems 3 and Corollary 3.1. We start by obtaining new results about those schemes that depend on majority votes over pairs of alternatives (see the expository article by Nierni and Riker (1976).) For example, an alternative is called a ***Condorcet winner*** it is receives a majority vote when compared against each of the other alternatives. A Condorcet winner does not always exist (e.g., the above, second example), so other schemes have been proposed to determine the winning alternative. The following definition appears to include all methods based on the ordinal rankings.

Definition

A ***binary ranking method*** is a nonconstant mapping from a subset of BR into $\{a_1, \ldots, a_N\}$. That is, based on the ordinal rankings of pairs of alternatives, one of the N alternatives is selected.

Examples

(1) A Condorcet winner is a binary ranking method. The subset is the set of all elements of BR where some one alternative is preferred to all other alternatives.

(2) An obvious extension of the Condorcet winner is to select the alternative that wins the largest number of pairwise comparisons. This extension admits a larger subset of elements from BR.

(3) Suppose $N - 1$ alternatives are proposed to replace the status quo, a_1. The selected alternative is a_1 if and only if a_1 is a Condorcet winner. If not, then from the set of those alternatives that beat a_1, select the one that wins the most pairwise comparisons.

(4) A commonly used binary ranking method is an agenda.

Definition

Let $N > 3$. An agenda is an ordered listing of the N alternatives. The first two listed alternatives are voted upon. The alternative receiving the majority vote is then compared with the third listed alternative. This iterative, pairwise comparison procedure is continued to the end of the listing. The remaining alternative is the selected alternative.

The following statement extends several results from "agenda manipulation" (see, for example, McKelvey (1976) and Plott and Levine (1978)). It implies that the right to set an agenda for a meeting is a potential source of power (the first conclusion) that may lead to an undesired outcome (the second conclusion).

Corollary 3.2

Let $N > 3$. There exist vectors' profiles and N agendas such that, when the *same* voters use the j^{th} agenda, the outcome is $a_j, j = 1,\ldots,N$. For $N > 3$, there exist voters' profiles and N agenda so that the above holds even though ***all the voters*** prefer a_3 *{to* a_4, a_4 *{to}* $a_5,\ldots,$ *{and}* a_{N-1} *to* a_N.

An interesting feature of this corollary is that for a fixed profile of voters, the winning alternative varies over all possible outcomes as the "seeding", or the choice of the agenda, changes. The proof depends on the fact that majority, pairwise voting can define any desired cycle and subcycle. Thus, this conclusion extends to the other binary ranking methods that depend on the initial seeding. This includes tournaments, whether single, double, or k-fold elimination, certain hierarchical methods, etc. In particular, because a change in the seeding changes the definition of a binary ranking method, Corollary 3.2 is a special case of the following.

Corollary 3.3

Let P_1 and P_2 be different mappings from BR to $\{a_l,...,a_N\}$. There exist profiles of voters so that the outcome of the two binary voting methods differ.

The next result compares the outcome of a binary ranking method with the election results of a weighted voting system. As special cases, it shows that a Condorcet winner, or the result of an agenda, need not agree with an election ranking.

Corollary 3.4

Let $N > 3$. Let the set Al_k be ranked with the voting vector $W_k, K = 3,...,N$. Let a binary ranking method be given. Assume that $W_k, k = 3,...,N$, is not reverse neutral. Let $Rk_k, k = 3,...,N$, by any ranking of Al_k, and let a_j be an arbitrary element in the image of the binary ranking method. There exist voters' profiles such that (i) the election result of $Al_k, K = 3,...,N$, and (ii) the binary ranking method selects a_j.

As a consequence of this theorem, there is a profile of voters so that their plurality election ranking of S_k is $a_1 > a_2 > ... > a_k$ if k is even and the reverse of this if k is odd, and the Condorcet winner is a_1.

This chaotic state of affairs cannot be eliminated if the selection method is defined to combine, in some way, the election results over all of subsets $Al_k, K = 2,...,N$. For instance, in a run-off election, the lower ranked alternatives are dropped, and the remaining set is reranked in a separate election. The following definition extends this notion.

Definition

A dynamical selection process consists of (i) a set of voting vectors $\{W_N,...,W_3,(1,0)\}$, (ii) rules that eliminate a specified, positive number of alternatives from a set of k alternatives, $k = 2,...,N$, and (iii) a selection function. The procedure is defined in the following way. The set Al_N is ranked by using W_N. Then, based on the elimination rule for the N alternatives, $N-s$ alternatives are eliminated. The remaining set of s alternatives is ranked by using W_s. Iteratively, this procedure is continued. Based on these election rankings, the nonconstant selection procedure selects one alternative.

Examples

(1) The standard "run-off" election is a dynamical procedure. At each step, the bottom ranked alternative is eliminated. The selected alternative is the one remaining at the end of the process.

(2) This run-off procedure can be generalized in the following way. Choose a positive integer $k < N$. The elimination procedure is the same as in (1), but the selection rule selects the top ranked alternative from the election ranking of k alternatives. If $k = 2$, this is the above procedure. If $k = N$, this is a standard election procedure.

(3) The run-off procedure can eliminate more than one alternative at each stage. For instance, after the N alternatives are ranked, all but the top two alternatives may be dropped.

(4) Let a_1 represent the status quo, and let $a_j, j = 2,\ldots,N$ represent the contending alternatives. Use W_N to rank the N alternatives. If a_1 is the top ranked alternative, it is declared the winner. If it is not, then eliminate a_1 and rank the remaining alternatives with W_{N-1}. The top ranked alternative from this election is declared the winner.

(5) The elimination rule may depend on the alternatives. For instance, the process described in (4) can be modified to eliminate not only a_1 but also all alternatives ranked below a_1 in the first election.

As a special case, the following asserts that the winner of a run-off election need not be a Condorcet winner.

Corollary 3.5

Let $N > 3$. Assume that a binary ranking method and a dynamical selection process are given. Suppose that, for each $K \geq 3$, the weight vector W_k is not reverse neutral. Arbitrarily select a_j from the range of the binary ranking method and a_k from the range of the dynamical process. There exist profiles of voters so that the binary ranking outcome is a_j while the dynamical method outcome is a_k.

Recently there has been interest in election procedures where a voter can choose a voting vector to tally his ballot.

Definition

A *multiple voting system* used to rank N alternatives is determined by a set M_N of voting vectors where at least two of these vectors and E_N are linearly independent. Each vector ranks the N alternatives on his ballot, and then he selects a vector from M_N to tally his ballot.

Examples

(1) Bullet voting: The defining set of voting vectors is $M_N = \{(2,0,\ldots,0),(1,\ldots,0),\ (1,1,0,\ldots,0)\}$. This procedure was used during the 1970s for some legislative offices in Illinois.

(2) Cardinal voting: The set M_N contains all voting vectors where the components sum to unity. Occasionally, cardinal voting is used to define rankings for methods from decision analysis.

(3) Approval voting: The defining set of $N-1$ vectors is $\{(1,0,\ldots,0),(1,1,0,\ldots,0),\ldots,\ (1,1,\ldots,1,0)\}$. For this method, which was introduced by R. Weber, among others, and ithas been analyzed by Brams and Fishburn (1982), a voter indicates either approval or disapproval of each alternative.

Often the results of multiple voting systems are compared with the Condorcet winner. The following shows that the results can be incompatible.

Corollary 3.6

Let $N \geq 3$. Let a binary ranking method be given. Let M_k define a multiple voting system for $Al_k, k = 3,\ldots,N$. Assume that, for each k, M_k contains at least one vector that is not reverse neutral. Let Rk_k be a ranking for Al_k, and let a_j be an alternative in the range of the binary ranking method. There exist voters' profiles so that (i) the multiple election result for Al_k is $Rk_k, k = 3,\ldots,N$, and (ii) the binary ranking outcome is a_j.

A consequence of this result is that, for any choice of s, there exist examples where the alternative ranked in s^{th} place in an approval voting election is the Condorcet winner, and there exist examples where the results based upon approval voting are $a_4 > a_3 > a_2 > a_1$ for the set of four alternatives, $a_1 > a_2 > a_3$ for the subset of three alternatives, and a_2 is a Condorcet winner.

It follows from this approach that the *principal cause of the social choice paradoxes is the difference between the dimensions of the domain and the range* of a comparison mapping (see Section 2, comment (2)). To model a weighted election for N alternatives, the domain $Si(N!)$ has dimension $N!-1$. The image is $E(A)$. Because the domain is $Si(N!)$, this image is in the simplex $Si(N)$ in R^N. (The sum of the components of W^N define this simplex. Without loss of generality, assume that this sum is unity.) Thus, the range space has dimension $N-1$. This difference of $N\{(N-1)!-1\}$ is zero if and only if there are only two alteratives. Therefore, if $N \geq 3$, other relationships with resulting paradoxes can be added. This is illustrated by Theorem 3.

This dimensional argument also proves that, for $N \geq 4$, Theorem 3 is *not* the "best possible" result. To see this, we need to describe the comparison mapping L for Theorem 3. The first N components of L are given by $E(A)$, the next $N-1$ by the expected value of the weighted voting method defined by W_{N-1}, etc. The last $N(N-1)/2$ components are given by the expressions $P(A_k) - P(A_j), k < j$. Thus, the range space for L is

$$Si(N) \times Si(N-1) \times \ldots \times S(3) \times J^{N(N-1)/2}$$

where J is the interval $[-1,1]$. This range space has the dimension $N + (N-1) + \ldots + 3 + \{(N)(N-1)/2\} = N^2 - N - 1$. For the model described in the theorem, the difference between the dimensions of the domain and the range is $N! - N^2 + N$. This value is positive if and only if $N \geq 4$.

Corollary 3.7

Let $N \geq 4$. In addition to the subsets of alternatives described in Theorem 3, $N! - N^2 + N$ additional relationships involving the rankings of the N alternatives can be defined in such a way that, for certain profiles of voters, the results are independent of the rankings obtained in Theorem 3.

A simple dimensional argument shows that even if Theorem 3 can be extended from nested sets of three or more alternatives to all possible subsets of alternatives, for $N \geq 4$ additional relationships can still be found.

To complete our description of L, notice that the comparison value on each simplex in the range is the point of complete indifference $N^{-1}E_N$. For each of the intervals J, the comparison value is 0. Thus, the comparison point is

$$(N^{-1}E_N,(N-1)^{-1}E_{N-1},\ldots,3^{-1}E_3,0,0,\ldots,0)\ .$$

Therefore Theorem 3 is an example of Theorem 2 where M and N are manifolds and where the comparison point is not the origin of a Euclidean space.

The comparison mapping L is linear. Therefore $L^{-1}(O)$ must be a linear subspace with dimension $N! - N^2 + N$. If $N > 3$ this space must intersect the boundary of $Si(N!)$. A boundary point of $Si(N!)$ corresponds to a profile of voters where none of the voters rank the alternatives in certain ways. This extreme boundary behavior describes the limits of the voting paradox. A special case is described in Corollary 3.2.

Corollary 3.8

Let $N \geq 4$. The results in Theorem 3 can be obtained with profiles of voters where no voter has certain rankings of the alternatives.

We have not tried to find a general characterization of the boundary behavior.

The proof of Theorem 3 involves showing that the linear comparison map has maximal rank. The rank condition does not hold if $W_k, k = 3,\ldots,N$ are Borda vectors. It turns out that L has corank (with respect to the range space) of at least $\{N(N-1)/2\} - 1$. This means that although a Borda election ranking admits inconsistencies with respect to a given binary ranking method, not all possible inconsistencies are admitted. In particular, *Theorem 3 does not hold if even one of the* W_k is a Borda vector. A direct verification of this for $N = 3$ is given in Section 4.

With only slight modifications, these results can be used to describe certain ranking procedures coming from probability and statistics. For instance, suppose N forms are making the same product, and they are to be ranked based on the quality of their products. In Theorem 3, identify the "i^{th} alternative" with the "i^{th} firm" the "t^{th} voter" with the "j^{th} vector sample" of the product taken from each of the N firms, and the "j^{th} voter's preference ranking" with the linear "quality ranking" of the products in the j^{th} sample. The relationship $a_k > a_j$ means that, based upon the samples, form k's product appears to be superior to firm j's. It follows from Corollary 3.1 that any possible choice of binary rankings is realized by an open set of data points. Binary sampling approaches need not lead to a linear ordering of the "quality of the firms". Indeed, in this way, the well-known Steinhaus-Trybula paradox (Steinhaus and Trybula, 1959), where the final ranking of three forms is $a_1 > a_2, a_2 > a_3$, but $a_3 > a_1$, becomes a special case of Corollary 3.1.

It follows from Theorem 3 that, even if the firms are ranked by use of weighted ranking methods, the results could be difficult to interpret. For instance, the weight vectors W_k = (1,0,...,0) correspond to the natural ranking method based on $P(X_i = \max\{X_j : j \in A_k\})$. It follows from the above that, should some one firm be deleted, the revised ranking could drastically change. Other measures experience similar problems. A similar effect occurs for the scoring of athletic events where a voter's ranking corresponds to how the various teams are placed in a particular event, etc.

As a final amusing example, note that a connoisseur is often described as a person whose taste preferences are based upon several attributes (e.g., the color, the taste, and the bouquet of a wine), and whose rankings are based on an aggregation of them. If so, we should not expect his binary comparisons to define a transitive ordering. This is, of course, an N alternative version of the famous folklore "pie" example (I prefer "apple" to "cherry", but if "blueberry" is available, then my choice is "cherry").

As in Section 2, there exist open sets in the domain which exhibit each of the above behaviors. Consequently, these examples cannot be dismissed as being isolated; the behavior is robust. As the number of agents increases (the denominators of the rational points become larger), so do the number of the possible examples, which leads us to the following corollary.

Corollary 3.9

Consider a system of weighted voting methods as described in Theorem 3. Let Q denote an outcome over the various sets as described in Theorem 3. Let $n(Q,m)$ be the probability that the election result for a group of m voters is Q. Assume that the profiles of voters are uniformly distributed. Then, as $m \to \infty, n(Q,m)$ approaches the ratio of the area of $L^{-1}(Q)$ to the area of the simplex $Si(N!)$.

For elementary number theoretic reasons, the sequence $\{n(Q,m)\}$ may not be monotone. The limit is positive if $L^{-1}(Q)$ contains an open set; this is true whenever Q does not admit ties. For other distributions, the ratio is determined in a similar fashion, but with a different measure.

4. PROOFS

The proof of Theorem 2 is obvious. To prove Theorem 3, we first prove Corollary 3.1.

Proof of Corollary 3.1

List the pairs of alternatives in the following order: the first pair is (a_1, a_2), the second set of two pairs is given by $(a_j, a_3), j = 1,2$, and the j^{th} set of k pairs is given by $(a_j, a_{k+1}), j = 1,\dots,k;\ k = 3,\dots,N-1$. A ranking of the N alternatives defines an $\{N(N-1)/2\}$-dimensional vector in the following way. The j^{th} component is determined by the ranking of the j^{th} pair of alternatives. This component is 1 if the first listed alternative is preferred to the second; otherwise, it is -1. For example, the vector associated with the preference ranking $a_1 > a_2 > \dots > a_N$ has the value 1 in all the components.

Because the N alternatives can be ranked in $N!$ different ways, the comparison map is a linear mapping from $Si(N!)$ to $J^{N(N-1)/2}$ where J is the interval [-1,1] and the comparison point is O. (this map defines a convex combination of the above $N!$ vectors.) We must show that there is a point p in the interior of $Si(N!)$ such that (i) p is in the preimage of O and (ii) the Jacobian of the comparison map at p has full rank. Let $p = (N!)^{-1}(1,\dots,1)$. Because p is the profile where there are equal numbers of voters for each of the $N!$ possible ways to rank the alternatives, p is mapped to O.

The comparison mapping is linear, so it has a matrix representation. The matrix is the Jacobian, and it consists of the $N!$ column vectors defined above. It remains to show that this set of $N!$ vectors includes $N(N-1)/2$ linearly independent vectors.

Consider the vectors $V_j, j = 1,\dots,N(N-1)/2$, where V_j has the value 1 for the first $\{N(N-1)/2\} - (j-1)$ component and -1 for the remaining components. This set of vectors is linearly independent. This is because they form a square array where the entries on and above the diagonal from the lower left-hand corner to the upper tight-hand corner are all equal to 1. All the other entries are -1.

There are 2^N vectors with entries of either 1 or -1. Most of them are not related to the described ranking method. So, to complete the proof, it remains to show that each V_j is associated with one of the $N!$ rankings of the alternatives. The choice of the components and the vectors V_j makes this fairly simple. The vector V_1 corresponds to the ranking $a_1 > a_2 > \dots > a_N$. Vector V_2 has -1 only in the last component; this corresponds to a transposition of a_N and a_{N-1}. These two alternatives are adjacent in the first ranking, so the ranking for V_2 can be obtained from the ranking for V_1 by transposing these alternatives. This defines the ranking $a_1 > a_2 > \ldots > a_{N-2} > a_N > a_{N-1}$.

Indeed, the only difference between V_j and V_{j+1} is in one component. This component reflects a change in the ranking of precisely one pair of alternatives. By construction, these two alternatives are adjacent in the ranking Rk_j that is associated with V_j. Therefore, the ranking for V_{j+1} is obtained by transposing these two adjacent alternatives in Rk_j. This completes the proof.

This proof is based on the fact that the -1's in the square array correspond to the $N-1$ adjacent transpositions required to move a_N from last place in $a_1 > a_2 > \ldots > a_N$ to first. This defines $N-1$ rankings where the last one is $a_N > a_1 > a_2 > \ldots > a_{N-1}$. Next, move a_{N-1} from what is now last place to second, etc.

Proof of Theorem 3

Let the weight vectors $W_k, k = 3,\ldots,N$, be as specified in the statement of the theorem. With each ranking of the N alternatives, we associate a vector with $\{N(N-1)/2\} + 3 + \ldots + N$ components. The first $N(N-1)/2$ components are defined as above. The next three are given by the appropriate permutation of W_a to correspond to the specified ranking. For instance, the ranking $a_2 > a_3 > a_1$ is identified with the vector (w_3, w_1, w_2). In general, the set of k components is the appropriate permutation of W_k to reflect the ranking of the k alternatives, $k = 3,\ldots,N$. The comparison mapping L which is a mapping from $Si(N!)$ to $J^{n(N-1)/2} \times Si(3) \times \ldots \times Si(N)$ is described in Section 3. The point p described above is mapped to the comparison point $(O, O, \ldots, O; (1/3)E_a, \ldots, (1/N)E_N)$.

Because L is linear, its matrix representation defines the Jacobian. This matrix has $N!$ column vectors with $\{N(N-1)/2\} + 3 + \ldots + N = N^2 - 3$ components. (The dimension of the range space is smaller; it has dimension $N^2 - N - 1$. The difference results from the constraints defining the $N-2$ simplices $Si(k)$ in the image space.) We must show that there are $N^2 - 3$ linearly independent vectors.

In the proof of the corollary, a set of $N(N-1)/2$ vectors that are independent in the first $N(N-1)/2$ components were found. To obtain the remaining $\{N(N+1)/2\} - 3$ independent vectors, take the vector associated with each of the $N!$

rankings and add it to the vector associated with the reversal of this ranking. Each of the first $N(N-1)/2$ components of the vector associated with the reversed ranking will differ in sign from the original vector. Therefore, the sum vectors will have zeros in each of these first $N(N-1)/2$ components. Consequently, these new vectors are orthogonal to the range space used in the proof of Corollary 3.1. All we need to do is to show that these new vectors contain a set of $\{N(N+1)/2\}-3$ independent vectors.

For a reverse neutral vector, these new vectors are all multiplies of E_N. In all other cases, the j^{th} component has the value $w_j + w_{N-j+1}$. For instance, for the voting vector (4, 3, 0), the new vector corresponding to the rankings $a > b > c$ and $c > b > a$ is (4, 6, 4). The vector corresponding to $a > c > b$ an $b > c > a$ is (4, 4, 6). In general, these new vectors would correspond to a voting vector except that they do not satisfy the monotonicity condition. However, the results given by Saari (1984) hold even for vectors that do not satisfy these monotonicity properties. Therefore, the above reduces to a special case of the one given by Saari (1984). This completes the proof.

Proof that a Borda weight vector does not work for $N = 3$

Assume that the alternatives are a, b, and c. assume that the Borda weight vector is $B_3 = (3,2,1)$. The comparison mapping is linear and its image includes the comparison point (0,0,0;6,6,6). (B_3 is not normalized, and so the sum of the components of $Si(3)$ is 6.)

The comparison regions in $Si(3)$ are identified with the linear rankings of the three alternatives. To obtain them, note that if the axes of R^3 are labeled in the usual x,y,z notation, then the region $x > y$ corresponds to $a > b$, $y > z$ corresponds to $b > c$, $z > x$ corresponds to $c > a$, etc. In this way, the simplex $Si(3)$ is divided into six open sets which are defined by the intersection of the simplex with the three hyperplanes $x = y, y = z$, and $z = x$ (see Saari (1978, 1982)).

Suppose that the different behaviors described in the theorem hold for B_3. This means that the image of the comparison mapping meets each of the six regions of $Si(3)$ as well as all of the open regions in J^3. In all, it would meet 48 open regions. If this happens, then, by the linearity of the mapping and a comparison of the dimensions of the domain and range, it follows that the mapping is onto a neighborhood of the comparison point. This forces the matrix to be of rank five. To show that this is not so, we list all six of the vectors and then extract a four-dimensional basis.

The vectors are as follows:

$a > b > c$, (1,1,1;3,2,1); $a > c > b$, (1,1,−1;3,1,2) ;

$c > a > b$,(1,−1,−1;2,1,3); $c > b > a$, (−1,−1,−1;1,2,3);

$b > c > a$, (−1,−1,−1;1,3,2);$b > a > c$, (−1,1,1;2,3,1)

These six vectors admit a basis consisting of the first three vectors and the vector (0,0,0; 1,1,1). Thus, the system has corank 2. Since this last vector is orthogonal to the image space, the system has corank 2 with respect to the image space. This, and the linearity of the mapping, means that the comparison mapping has a nonzero intersection with 12 of the 48 admissible comparison regions. If the mapping were always consistent, then the mapping would meet only 3! = 6 regions. Thus, the mapping still admits several "inconsistent" conclusions. (In a paper being prepared, we characterize the election rankings admitted by a Borda count.)

Extension of Theorem 3

The last part of the proof of Theorem 3 is based on the work of Saari (1984) which admits a wider variety of results. For example, for k alternatives, suppose there are $k-1$ weight vectors W_k, which form, with E_k, a linearly independent set. Arbitrarily choose $k-1$ rankings of the k alternatives. The theorem asserts that there exist voters' profiles so that when the same voters rank the set of k alternatives, $k = 3,...,N$ with the i^{th} voting vector, then the outcome is the i^{th} specified ranking of the alternatives. This is true for all choices of $i = 1,...,k-1$ and k.

A similar extension holds for Theorem 3. For each $k = 3,...,N$, choose $k - 1 - k/2$ voting vectors with the following property: (i) the voting vector is not reverse neutral, and (ii) the set of $k - k/2$ vectors, defined by E_k and the vectors formed by adding each of the $k - 1 - k/2$ voting vectors to its reversal, is a linearly independent set. For each voting vector, arbitrarily choose a ranking of the k alternatives. For each pair of alternatives, designate one of them. There exist voters' profiles so that when the i^{th} voting vector is used to rank the k alternatives the outcome is the assigned ranking $i = 1,...,k - 1[/2]$, $k = 3,...,N$. For each pair, a majority of the same voters prefer the designated alternative. The proof of this statement is a straightforward modification of the proof of Theorem 3.

Proof of Corollary 3.2

From Corollary 3.1, it follows that there is an open set of voters' profiles where the outcome in pairwise elections is the cycle

$$a_1 > a_2, a_2 > a_3, \ldots, a_{N-1} > a_N, a_N > a_1 \quad .$$

Consider the reversed cycle $a_1 < a_N < a_{N-1} < \ldots < a_2 < a_1$. The following defines an agenda where a_j will be the winner. Let the agenda be the N terms in the reversed cycle that start with the alternative immediately following a_j and ends with a_j. For example, a_3 wins with the agenda $[a_2, a_1, a_N, a_{n-1}, \ldots, a_4, a_3]$.

The second part of this corollary is a consequence of the boundary properties of $L^{-1}(O)$. The second part of this corollary is a consequence of the boundary properties of $L^{-1}(O)$. The profile, where an equal number of voters have each of the three rankings $a_1 > a_2 > \ldots > a_N, a_2 > a_3 > \ldots > a_N > a_1$, and $a_3 > a_4 > \ldots > a_N > a_1 > a_2$ has the desired properties. Note that in each pairwise comparison the winning alternative receives either $\frac{2}{3}$ of the vote, or all of it! This is true for whichever agenda is used and whichever alternative wins.

The dice example

This is a straightforward computation. However, the domain point used in the image of the comparison point should correspond to two identical weighted, but not fair, dice. The probability that a particular face will surface is left to the end of the computation. In other words, there are some complications in the computation with two fair dice.

ACKNOWLEDGEMENT

This research was supported by NSF Grants IST-8111122 and IST-8415348. I am pleased to acknowledge conversations with W.W. Funkenbusch, Tom Miles, and Hans Weinberger on these and related topics. Some of this work was done while I was visiting the Institute for Mathematics and its Applications (IMA) at the University of Minnesota. I would like to thank my hosts, Hans Weinberger, George Sell, and Leo Hurwicz for their kind hospitality during my stay. These results were presented at an IMA conference on mathematical political science in July 1984.

REFERENCES

Blyth, C. (1972a). On Simpson's paradox and the sure-thing principle. *J. Am. Stat. Assoc.* 67, 364-366.

Blyth, C. (1972b). Some probability paradoxes in choice from among random alternatives. *J. Am. Stat. Assoc.* 67, 366-373.

Brams, S., and Fishburn P. (1982). *Approval Voting*. Birkhauser, Boston.

Fishburn, P. (1981). Inverted orders for monotone scoring rules. *Discrete Appl. Math.* 3, 27-36.

McKelvey, R.D. (1976). Intransitivities in multidimensional voting models and some implications for agenda control. *J. Econ. Theory 12, 472-482.*

Niemi, R., and Riker W. (June 1976). The choice of voting systems. *Sci. Am.* 234, 21-27.

Plott, C.R., and Levine M.E. (1978). a model of agenda influence on committee decisions. *Am. Econ. Rev.* 68 146-160.

Saari, D.G. (1982). Inconsistencies of weighted summation voting systems. *Math. Oper. Res.* 7, 479-490.

Saari, D.G. (1984). The ultimate of chaos in weighted voting systems. *Adv. Appl. Math.* 5, 286-308.

Saari, D.G. (October 1978). *The Geometry of Departmental Politics, the Scoring of Track Meets, and Arrow's Social Choice Theorem.* Northwestern University preprint.

Steinhaus, H., and Trybula S. (1959). On a paradox in applied probability. *Bull. Acad. Polo Sci.*, 7, 67-69.

Vaupal, J.W., and Yashin, A.I. (1985). Heterogeneity's ruses: Some surprising effects of selection on population dynamics, *The American Statistican*, 39, 176-185.

Wagner, C. (1982). Simpson's paradox in real life, *The American Statistican*, 36, 46-48.