Discussion Paper No. 712

SYMMETRY, VOTING, AND SOCIAL CHOICE

by

Donald G. Saari
Department of Mathematics
Northwestern University
Evanston, IL 60201

November 1986

Eventually I'll outline how symmetry arguments, as captured by the wreath product of permutation groups, help to resolve several long standing research problems. To introduce the issues, consider why chairing a department can be a "hair losing" proposition. Start with the spring banquet where only one beverage can be served. In a hypothetical 15 member department, 6 specified to the Chair that they preferred milk to wine to beer (milk>wine>beer), 5 specified beer>wine>milk, and 4 specified wine>beer>milk. The Chair's decision was easy; the group's plurality ranking is milk>beer>wine with a tally of 6:5:4. Accordingly, the Chair announced that milk would be served. For unexplained reasons, milk wasn't available; so he ordered the Department's second choice - beer.

During the banquet, the naturally inquisitive wine lovers discovered that beer <u>wasn't</u> the Department's second choice; 2/3 of the department (10 to 5) preferred wine to beer! With aroused suspicions, further investigations proved that 3/5 of the faculty (9 to 6) preferred wine to milk and 3/5 preferred beer to milk! Rumors floated that the Chair reversed the Department's "true" ranking, wine>beer>milk, so the outcome would favor his aggravated ulcer.

Once impugned, suspicions grow. The unrest resulted in a departmental meeting to decide among the competing proposals: 1: "The Chair is commended for his efforts." 2: "To help the Chair tally ballots, he is to teach remedial math." 3:"The Dean must replace the Chair." After a lively corridor debate, the department split evenly among the choices 1>2>3, 2>3>1, and 3>1>2. Our Chair tried to protect himself by exploiting his right to set the agenda. His strategy was to focus debate on his preferred alternatives 1 and 2 by having the first vote between them; then, the winning proposal would be matched against 3. It failed; the dreaded 3 was overwhelmingly adopted. (Both votes were by landslide tallies of 10 to 5.)

What a missed opportunity! Had the Chair used the agenda that paired the winner of 2 vs.3 with 1, he would have left the meeting with "proof" that his problems were caused by a small dissident minority. With this agenda, 2 would have beaten the feared 3, and his preferred 1 would have won in the second vote. Both tallies would have been decisive - 10 to 5.

The Dean already had her doubts about the Chair's integrity. Earlier, she asked him to choose a calculus book to better prepare the students for Physics. In response, he divided the students into two groups of 300 each where one group used the new Sorry book and the other used the standard Pathetic text. The choice was to be based on how the students did on a Physics exam at the end of the term. According to the occasionally reliable student newspaper, both on North and on South Campuses a higher percentage of the Sorry students passed the exam than the Pathetic students. But, the Chair asserted that the Pathetic text did better! (On North Campus, 90 out of 240 Sorry students passed compared to 20 out of 60 Pathetic students. On South campus, the results were 30 out of 60 compared to 110 out of 240. Thus, only 120/300 of the Sorry students passed compared to 130/300 of the Pathetic students.)

Action had to be taken. After consulting with the Political Science Department about procedures, the Dean announced a Departmental election to be tallied according to a method proposed in 1781 by the eminent French mathematician J-C.Borda. This is where, with N alternatives, N-i points are assigned to a voter's i^{TH} ranked candidate. Back in the department, a coffee room survey showed that 7 preferred Abbott>Boyce>Chair, 7 preferred Boyce>Abbott>Chair, and our beleaguered Chair preferred Chair>Abbott>Boyce. In this de facto two person race, Boyce would lose. Consequently, Boyce's supporters acted "strategically" by voting Boyce>Chair>Abbott so that each of them would give Boyce a two point, rather than a single point differential over Abbott. Of course, Abbott's supporters suspected this and marked their ballots Abbott>Chair>Boyce. The Chair was reelected.

2. Social Choice

Admittedly, this is a farfetched story, but hopefully it created doubt in your mind about how your department filled that one tenure track slot, about your last vote for the Chair, or about the election results for your Departmental Budget Committee. It should, because this tale illustrates some very real voting problems. We tend to dismiss the choice of a tallying method as being a minor issue, but it isn't. Different methods can lead to completely different election outcomes. For instance, with our beverage example, the Borda Count outcome is wine>beer>milk with the tally 19:14:12 - this is the exact reversal of the plurality ranking!

What else can happen? What is the "best" choice? These are the types of questions raised in the 1780's by the French mathematicians Borda and Marquis de Condorcet. One of Borda's main contributions was his tallying procedure. But his critics of that time, including Laplace, discovered several weaknesses. For instance, as illustrated above, it's easy to try to manipulate his system. (This still serves as an argument against using it.) A more serious criticism concerns the lack of any justification for choosing these particular weights. Why not assign 5 points to a first place candidate, 3 to a second, and zero to all others? How about 4, 3.9, and 0? Indeed, there is a vector space of possibilities. If we define the <u>voting vector</u> $\underline{W}=(w_1,w_2,..,w_N)$ to consist of the weights where w_j points are assigned to a voter's j^{TH} ranked alternative, $j=1,..,N$, then the only constraints are that $w_j \geq w_{j+1}$, $j=1,..,N-1$, and $w_1 > w_N \geq 0$. What justifies using $\underline{B}=(N-1,N-2,..,0)$?

In contrast, Condorcet believed the emphasis should be on the pairs of alternatives. He argued that if an alternative always wins by a majority vote when compared with any other alternative, then it should be adopted. (Wine and Abbott are Condorcet winners.) If you accept Concorcet's views, then Borda's method has another flaw; it doesn't always elect a Condorcet winner.

The modern theory of of Social Choice doesn't provide much help for this type of question. This theory dates to K. Arrow's book [1] where he proved it is impossible to construct a method for $N \geq 3$ alternatives that satisfies certain simple, desirable, and seemingly innocuous axioms. His unexpected, shocking conclusions and his axiomatic, combinatoric approach set off an avalanche of papers. (See, for example, [2].) But, it is only a slight exaggeration to say that almost all of these modern results are negative: starting with some desirable properties, combinatorics are used to create examples proving that only those systems such as dicatorships, oligarchies, etc., can satisfy them. These frustrating conclusions provide very little guidance in the choice of a system. Indeed, they only heighten the mystery; what's the source of the difficulties? (A sneak preview: the wreath product of permutation groups.)

To address these issues, I'll start with tallying methods (voting vectors). Ideally, we'd like to discover everything that could possibly go wrong. Namely, we'd like to have a catalog of all possible election outcomes over all subsets of candidates for all possible choices of voting vectors. To make sense of this, note that N alternatives determine 2^N subsets of candidates where one subset is empty and N of them have only one element. This leaves $2^N-(N+1)$ subsets with enough alternatives (at least two) to be ranked by an election. Label these subsets in some order, let $\underline{W}_J$ be the voting vector assigned to tally the ballots for the j^{TH} subset, and let $\underline{W}^N=(\underline{W}_1,..,\underline{W}_{2^N-(N+1)})$ be the listing of these voting vectors.

Suppose a profile of voters, $\underline{p}$, is specified. (A profile is a listing of each voter's ranking of candidates; we impose no restriction on number of voters.) Once $\underline{p}$ is given, $\underline{W}^N$ uniquely determines the election outcomes for these same voters for each of the $2^N-(N+1)$ subsets. Call the resulting sequence of election rankings, denoted by $f(\underline{p},\underline{W}^N)$, a <u>word</u>, and call the set of all possible words, $D^N{}_W=\{f(\underline{p},\underline{W}^N) | \text{ for all } \underline{p}\}$, the <u>dictionary</u> generated by $\underline{W}^N$. For instance, in our beverage example, (wine>beer; wine>milk; beer>milk; milk>beer>wine) is just one word (determined by the specified profile) in the dictionary generated by voting vectors {(1,0),(1,0),(1,0),(1,0,0)}. If we could characterize all entries in all dictionaries for all choices of $\underline{W}^N$, we would have our catalog of all faults and paradoxes admitted by tallying procedures.

This I've done. [3] To describe the results, the dictionaries need to be viewed as subsets of a universal set. So, let S^K be the set of all possible complete, binary, transitive rankings of the alternatives in the K^{TH} subset of alternatives and let U^N, the universal set, be the product of these $2^N-(N+1)$ sets. In other words, U^N contains all possible listings of election rankings, over all subsets, whether or not they make sense. Of course, we hope that $D^N{}_W$ is a small subset of U^N.

Theorem. If each of the $2^N-(N+1)$ voting vectors defining $\underline{W}^N$ correspond to a plurality election, then

1. $$D^N{}_W = U^N.$$

Indeed, except for a lower dimensional algebraic subset of $\underline{W}^N$'s, Eq. 1 holds.

Anything can happen! For almost all choices of voting vectors, any "paradox", any wild or weird listing of election outcomes actually can occur! As an extreme, use a random number generator to choose the election rankings for each of the $2^N-(N+1)$ sets of candidates. According to the theorem, there is a profile of voters so that their election ranking for each subset of candidates agrees with the generated result. Namely, the same voters vote sincerely over each of the $2^N-(N+1)$ subsets of candidates - they never change their minds - yet their election outcomes over the subsets of candidates coincide with the randomly generated rankings! This is most disturbing because the purpose of an election is to capture some type of consensus of the voters. It is difficult to accept that a voting procedure does this if the outcomes can be so indeterminate and depend so sensitively on which subsets of candidates just happen to be presented!

There are many consequences of this result, and in my partial sampling I'll "pick on" plurality voting only because it is so commonly used. (For almost any other choice of $\underline{W}^N$, the same statements hold.) The first one extends the beverage example (k=3) by illustrating that even though the voters' rankings of pairs is extremely well behaved, their election results over other sets of candidiates can vary, say, with multiples of three.

Corollary. There exists a profile of voters so that for each pair of candidates, a majority of the voters prefer a_j to a_k iff $j<k$. Nevertheless, for the same voters, their plurality election ranking of the sets $\{a_1,a_2,..,a_k\}$, $k=2,..,N$, is $a_k>a_{k-1}>...>a_1$ if $3|k$, $a_1>a_2>...>a_k$ if $3|k+1$, and $a_k>a_1>a_{k-1}>....$ if $3|k+2$.

In the beverage example, the plurality election not only relegated the Condorcet winner to last place, but it elected the Condorcet loser. The next statement illustrates that plurality elections can show even more disregard for Condorcet's winners and losers by doing this same thing for <u>all</u> subsets.

Corollary. There exists a profile of voters so that for each pair of candidates, $\{a_j,a_k\}$, the majority outcome is $a_j>a_k$ iff $j<k$. For the same voters, the plurality ranking for each subset with three or more alternatives is the transitive ranking generated by the <u>reversed</u> binary relationship $a_j<a_k$ iff $j<k$.

So much for any runoff election procedures based on plurality, or almost any other voting vector! The final outcome need not have anything to do with how the voters rank the pairs of alternatives, or, for that matter, how they rank almost any other subset of these candidates. To see this, choose rankings where a_1 is top ranked for <u>all</u> of the subsets containing it except for the set of all N alternatives. In this set, choose a ranking where a_1 is ranked just below the cutoff point that determines which candidates advance to the next stage of the runoff election. According to the theorem, a profile of voters can be found that simultaneously realizes all of these specified election outcomes. Arguably, a_1 is their favorite (it is much stronger than just being a Condorcet winner!), but a_1 doesn't even survive the first stage of the runoff election.

These examples may lead you to believe that Condorcet's approach is the correct one. I'm not so sure. By using Eq.1 and concentrating only on the pairs of candidates, it follows that the majority vote over pairs can define <u>any</u> conceivable binary relationship with cycles, subcycles, etc. Consequently, a Condorcet winner, or any obvious extension of this concept, need not always exist.

Our agenda example is another manifestation of the binary behavior. An <u>agenda</u> is a listing, say $[a_3, a_2, .., a_N]$, of the N alternatives. The first two, a_3, a_2, are voted upon, and the one receiving a majority vote is compared with the third listed alternative. This iterative process continues until only one alternative, the "winner", remains. As suggested by the "Departmental meeting", it's possible that the winner more accurately reflects the choice of the agenda rather than the preferences of the voters! Such "agenda manipulation" opportunites provide a savvy Chair with considerable, unintended power.

Corollary. There exists a profile of voters and N agendae so that when the j^{TH} agenda is used, the winner is a_j, $j=1,..,N$. Indeed, such profiles can be found so that in each of the the pairwise elections, the winning alternative receives <u>at least</u> 2/3 of the vote. In fact, if $N \geq 4$, all of this can occur even though <u>all</u> of the voters prefer a_j to a_{j+1} for $j=3,..,N-1$.

A profile illustrating the last statement is where the voters are evenly split among $a_1 \succ a_2 \succ ... \succ a_N$, $a_2 \succ .. \succ a_N \succ a_1$, $a_3 \succ .. \succ a_N \succ a_1 \succ a_2$.

Even though a_N isn't highly regarded by anyone, it wins with the agenda $[a_{N-1}, a_{N-2}, \ldots, a_1, a_N]$. (I'll leave it to the reader to find an agenda that selects a_j. The answer is suggestive of the role played by symmetry in the analysis of voting.) All sorts of other examples, such as showing how the outcome of various types of tournaments depend on the initial seeding, etc., can be created by using this theorem. I leave them to the interested reader.

Is there any good news? Is there a method where the election outcomes need not be so sensitive to which subset of candidates just happens to be presented? In other words, can a $\underline{W}^N$ be found where $D^N{}_W$ is a proper subset of U^N? Yes, and <u>the best choice is Borda's method</u>!

Definition. A voting vector $(w_1, \ldots, w_N)$ for N alternatives is a <u>Borda vector</u> iff $w_j - w_{j+1}$ is the same positive constant for $j=1,\ldots,N-1$. Let $\underline{B}^N$ denote a $\underline{W}^N$ where all of the voting vectors for sets of three or more candidates are Borda vectors, and let the dictionary be denoted by $D^N{}_B$.

It is a simple exercise to show that for a profile, the words generated by two different Borda vectors are the same. The following asserts that $\underline{B}^N$ minimizes the number and types of paradoxes.

Theorem. Let $N \geq 3$ and let $\underline{W}^N \neq \underline{B}^N$ be given. Then

3. $$D^N{}_B \subsetneq D^N{}_W.$$

In other words, if a profile defines a word demonstrating a flaw of the Borda Count, then all other choices of $\underline{W}^N$ admit this exact same word. Consequently, <u>any</u> criticism of Borda's method advanced by an example also serves as a criticism for all possible voting vectors! To illustrate, Borda's method need not elect a Condorcet winner, so this is true for all voting vectors. Conversely, any other $\underline{W}^N$ admits a large number of words (i.e., paradoxes) that aren't permitted by Borda's method. In fact, from our characterization of the entries of $D^N{}_W$ for any $\underline{W}^N$ (not given here)[1], it follows that Borda's method is the <u>only</u> one that doesn't elect a Condorcet loser nor rank a Condorcet winner in last place (e.g., see the Borda ranking of the beverage example), it is the <u>only</u> method that avoids many of the "runoff" election problems, it is the <u>only</u> method where $D^3_W \neq U^3$, the cardinality of D^N_B is <u>much</u> smaller than that of D^N_W for any $\underline{W}^N \neq \underline{B}^N$ (i.e., the Borda Count admits much fewer paradoxes)[2], etc. etc. So, if you want to minimize the number and the types of voting paradoxes, if you want to reduce the difficulty of interpreting what an election result "really means", then the unique choice is the Borda Count! This is an answer for Borda's critics.

3. Proofs via Statistical Paradoxes

What else can happen? Do these paradoxes reflect highly unlikely, specially constructed anomolies? With the above criteria, Borda's method is superior, but what about the telling criticism that it is obvious how to try to manipulate it? How about using methods where we can register our intensity of like or dislike for each candidate; e.g., how about giving each voter 10 points to distribute among the candidates in any desired way? The intuition for the answers of these and related issues comes from the proof of our theorem.

To outline the ideas, I'll use the "book selection" problem. This

----

1. The entries of a dictionary are "characterized" because there are too many to list; e.g., in general $|D^6_W| > 6.4 \times 10^{72}$. To appreciate this number, recall that the supercomputers are projected to do 10^{12} operations per second. Only 3×10^{27} seconds of time have elapsed since the "Big Bang". So, if a computer started at creation to list these outcomes, it would be about $1/10^{42}$ through. You may suspect from this number that for these theorems to be applicable, trillions of voters are required. But note, 6 candidates and 30 voters, generate $(6!)^{30} = 5.25 \times 10^{85}$ possible rankings. (The relevant number is smaller because of symmetries due to the "anonymity" of the voters, but this suggests why small numbers of voters can create so many different outcomes.) On the other hand, $(6!)^{15} = 7.2 \times 10^{42}$, so many of the paradoxes can't be realized with only 15 voters.

----

2. For instance, in general, $|D_{6B}|/|D_{6W}| < 1/10^{54}$.

doesn't concern voting; it illustrates what can happen when conditional probabilities are combined. Indeed, the book example is a special case of Simpson's Paradox, a paradox of particular concern in statistics because it describes what can occur when contingency tables are collapsed to determine the marginal information. The explanations of this peculiarity are based on the rules of combining conditional probabilities, so they don't address (nor are they intended to) what other, related phenomena can occur along with Simpson's Paradox. As we'll see, some rather surprising examples can be created.

The four relevant variables are determined by the two choices of books and the two sections on campus. Let x_j^C be the fraction of students passing the Physics exam that use the j^{TH} book, $j=1,2$, on campus C = NC (north campus), SC (south campus). Four more variables, d_j^C, determine the fraction of the students using each book on each campus. (So, $\underline{d}=(d_1^{NC},d_2^{NC},d_1^{SC},d_2^{SC})$ is on the unit simplex in 4 space, Si(4).)
The types of examples that occur are determined by the orthants of R^3 that meet the image of

4. $F:X=I^4 \times Si(4) \longrightarrow R^3$

defined by $(x_1^{NC}-x_2^{NC},x_1^{SC}-x_2^{SC},y_1-y_2)$,
$y_j=\Sigma(x_j^C d_j^C)/(\Sigma d_j^C)$.

Anything can happen if F meets all eight orthants or "ranking regions". This happens if at an interior point q in $X \cap F^{-1}(\underline{0})$, the Jacobian, DF, has maximal rank. (Because F maps an open neighborhood of q to an open neighborhood of $\underline{0}$, the image meets all ranking regions.) Indeed, from the continuity of F, each ranking region contains the image of an open set of X. This is important because finite examples are identified with the rational points in X. But, the rationals are dense, so the existence of these open sets implies there are an infinite number of finite examples illustrating each type of behavior. Moreover, because of these open sets, the examples cannot be dismissed as representing highly unlikely anomolies; they are robust.

This simple argument can be used to explain and extend many statistical and probabilistic paradoxes as well as our voting results. Furthermore, because all I'm using is an open mapping, an obvious modification of this argument determines what else can occur. To illustrate, I'll extend Simpson's Paradox by augmenting F with a mapping G to define $(F,G):X \to R^s$ where $(F,G)(q)=\underline{0}$, $s = \dim(X) = 7$, and the Jacobian of (F,G) at q is of rank s. Then, each orthant of the new ranking space, R^s, contains the (F,G) image of an open set of X. In this way, different choices of G create new paradoxes.

To figure out what choices of G can be used, notice that the three row vectors of DF define a three dimensional subspace. So, just by "eyeballing" the form of F, one can pick out what vectors aren't in this vector space. In the book example, such a vector could depend on all four variables describing North Campus, or all four variables describing South Campus. Consequently, any comparison (G) involving these variables, such as comparing expected waiting times for sampling without replacement, has a gradient vector of the appropriate form. For instance, let z_j^C be the probability that after two of the students using book j from campus C, C=NC, SC, are randomly selected (without replacement), at least one of them passed the Physics exam. If $G'=(z_1^{NC}-z_2^{NC}, z_1^{SC}-z_2^{SC})$, then (F,G') satisfies all but the dimension statement. This means a new extension of Simpson's Paradox can be created where on each campus, in two tries, it is less likely to find a Sorry student than a Pathetic one that passed the exam.

We can't extend G' to also create a waiting time paradox for the total school. This is because the first 5 comparisons force q to be in the set $x_1=x_2$, $x_3=x_4$, $d_1=d_2$, $d_3=d_4$, and the independence requires $x_1 \neq x_3$, $d_2 \neq d_3$. Thus, any new "paradoxes" cannot impose further restrictions on q. The maximal rank condition on $D(F,G')$ eliminates the possiblity of using "symmetric" comparisons of the form $g(x_1,x_2,d_1,d_2) - g(x_3,x_4,d_3,d_4)$. On the other hand, the last two comparisons could be of the form $(x_1-x_2)(2x_1+3x_3)$ and $(x_3-x_4)(5x_2d_3-x_4)$.

The same open mapping approach explains the paradoxes for other aggregation models, and, in particular, for voting, once an analogue for F is found. The N alternatives define N! types of voters according to the N! rankings without ties (or indifference). If p_j is the fraction of voters of the j^{TH} type, then p_j is one of the N! components of a vector $\underline{p}$ in Si(N!). The standard tally for an election is equivalent to a $\underline{p}$ tally just by dividing the tally by the total number of voters. So, $\underline{p}$ can be viewed as being a profile, and the domain for voting systems can be viewed as being the N!-1 dimensional simplex Si(N!).

The range is a product of simplices. A tally for a subset of candidates is a listing of candidates' individual tallies; this is a vector in an Euclidean space where each candidate's tally is identified with a coordinate axis. But $\underline{p} \in Si(N!)$, so the election tally is a vector in the simplex defined by the sum of the components of $\underline{W}_K$. Call this simplex Si_K, then

5. $f(-,\underline{W}^N):Si(N!) \text{---}\!> Si_1 x \ldots Si_{2^N-(N+1)}$.

The baricentric point of each simplex, which corresponds to a complete tie vote among all candidates, is a boundary point for all other ranking regions. So, it plays the same role as $\underline{0}$ in the book example. The replacement for the point q is a profile that causes a tie vote for all possible subsets. The obvious choice is where the voters are split equally among all possible types. Only the rank of Df is left, but this involves a symmetry argument, so I'll discuss it later.

4. From Multiple Systems to Strategic Voting

There are sorts of other paradoxes! This is because, for N>3, the dimension of the domain, Si(N!), is much larger than that of the range. So, as with the book example, the mapping $f(\underline{p},\underline{W}^N)$ can be augmented with other maps to create still wilder types of concomitant, unexpected behavior. (Some of them concern AMS and MAA procedures.) Many of these new paradoxes can be identified just by "eyeballing" the subspace generated by the row vectors of Df. For example, because this Jacobian depends on $\underline{W}^N$, other voting vectors need not be in this subspace. A consequence is our earlier assertion that election results tallied with different procedures need not agree. More precisely:

Theorem. To rank a set of N alternatives, let $(\underline{W}_{N1},..,\underline{W}_{NN-1})$ be N-1 voting vectors that, along with (1,1,..,1), span R^N. Choose N-1 rankings of the N alternatives. There exists a profile of voters so that when this profile is tallied with $\underline{W}_{NJ}$, the outcome is the j^{TH} selected ranking, j=1,..,N-1.

Thus, there is no reason to expect any consistency of election results among different tallying methods. For instance, the plurality ranking of the beverage example is the exact reversal of the Borda ranking. Examples are now easy to create; e.g., there is a profile of voters where their (1,0,0,0) tally is $a_1>a_2>a_3>a_4$, their (1,1,0,0) tally is $a_4>a_3>a_2>a_1$, while their (1,1,1,0) tally is $a_3>a_1>a_4>a_2$. Not much consistency here!

This result extends to all $2^N-(N+1)$ subsets to create a "super version" of our first theorem. (Again, the Borda Count minimizes possible outcomes.) As an illustration, there is a profile of voters so that their plurality ranking of $\{a_1,..,a_K\}$, is $a_1>a_2>..>a_K$ if k is even, and the reverse of this if k is odd, k=3,..,N. For the same voters and for each k, their (1,1,0,..,0) election ranking (vote for your top two candidates) always is the <u>exact reversal</u> of the plurality ranking.

All we need for such theorems is the independence of the augmenting map, so voting vectors can be replaced with other vectors, e.g., (0,0,..,0,1). In this way, we can analyze more sophisticated runoff procedures, such as the one used by the American Mathematical Society, where at each stage, candidates are dropped if they don't satisfy certain criteria. To see this, consider the simple system where at each stage the list of candidates is narrowed by dropping the candidate with the largest number of last place votes; i.e., the candidate with the largest (0,..,0,1) tally. An above type statement asserts there can be a lack of any relationship among outcomes; if anything other than a Borda Count is used to tally the election, then there are situations where the first candidate to be dropped is the Condorcet winner, etc. The beverage example illustrates another characteristic; the <u>winner</u> of the first stage, milk, is dropped!

An amusing paradox that emerges from this analysis is motivated by Simpson's Paradox where the aggregation of two favorable situations created an unfavorable one. Can this happen in voting? It can and in many different ways; I'll just identify one. It's obvious that positional voting methods are monotonic; if you vote, you're improving the chances of your candidates. On the other hand, our first theorem proves that monotonicity can be lost when results over different subsets of candidates are used, as in runoff elections. When a procedure involves several sets of candidates, it isn't difficult to show it need not be monotonic. Ah, this provides all sorts of opportunities such as showing that there are profiles where, by voting, the ballot <u>hurts</u> the voter's candidates; indeed, the vote has a reversed effect on the final outcome! In other words, there are situations where it is in a voter's best interest to <u>abstain</u>, even to <u>vote against</u> his or her candidates! (This is related to "strategic voting".) In fact, by using this approach, it isn't overly difficult to characterize these procedures. (Using completely different arguments, Steve Brams and Peter Fishburn [4] discovered one such procedure in terms of the standard runoff elections.)

I think I've made my point that the known and the previously undiscovered difficulties of tallying procedures (as well as in probability, nonparametric statistics, etc.) are consequences of the very large dimension of the domain. So, just imagine what new electoral mischief is created if the size of the domain is increased! This is an unexpected, unintended by-product of any "reform" procedure that provides voters with added options. For instance

Definition. A multiple voting system is an election procedure equivalent to having each voter specify his or her ranking of the N alternatives and then select a voting vector to tally this ballot from the set $M=\{(\underline{W}_{1N},..,\underline{W}_{jN}) \mid j\geq 2$, the difference between any two vectors is not a multiple of $(1,1,..,1)\}$.

Multiple voting systems are fairly common. The Mathematical Association of America recently adopted one - approval voting. This is where each voter indicates approval or disapproval of each candidate, so it is the multiple system defined by $M = \{(1,0,..,0), (1,1,0,..,0), .., (1,..,1,0)\}$. In my own department, we elect our Budget Committee by voting for no more than our top three candidates; namely, we use the system $M = \{(1,0,..,0), (1,1,0,..,0), (1,1,1,0,..,0)\}$. A commonly suggested multiple voting method is to let each voter distribute a fixed number of points among the candidates. If the voter has 10 points where any fraction of points can be assigned to any candidate, then M has an infinite number of voting vectors. Or, as used in some Illinois elections, only certain divisions of the points are allowed. Multiple systems are created whenever an organization is tolerant of "truncated ballots"; e.g., suppose for a Borda election, everyone is to rank the five candidates for a tenured position, but suppose some of the ballots list only one, or maybe two candidates. If all ballots are tallied, this creates the multiple system $M=\{(4,3,2,1,0), (4,0,..,0), (4,3,0,..,0)\}$.

A "reform" justification for using a multiple system is that it permits the voters to register the intensity of their likes and dislikes for the candidates. But, no good deed goes unpunished; the penalty accompanying this virtue forcefully comes from the significant increase in the dimension of the domain. In fact, the dimension is so large that it creates all sorts of new diabolical surprises for the tallying process. After all, not only must the domain represent the N! types of voters (the profile p), but also it must represent what fraction of each type of voter uses each of the possible voting vectors. This creates a fiber bundle over Si(N!) where a point in a fiber specifies the selection of the different tallying methods. If M admits K choices of voting vectors, then the dimension of a fiber is (K-1)(N!), so the dimension of the new domain is K(N!)-1. As just a sample of what new paradoxes can be admitted, Jill Van Newenhizen and I [5] showed that

Theorem. Let M be a given multiple voting system used to rank N>2 alternatives. There exists a profile of voters so that as the voters only change in their choice of how to tally the ballots, all N! election results occur.

Thus, multiple systems introduce complete indeterminacy even with the same set of candidates and with the same voters! You can see this with the beverage example; if approval voting is used in this election, then, as these same voters vary in whether they approve their first, or their first and second beverage, all 13 rankings result! (This includes the 7 rankings with ties.) Again, this statement extends to all of the subsets to create a general, multiple voting version of our basic theorem. By excluding an algebraic set that contains the Borda Vectors, Van Newenhizen and I showed that a profile can be found where the election rankings for each of these sets varies over all possible rankings as the voters changes choices of tallying procedures! As before, the results are robust because they are supported with open sets in the base space Si(N!).

This creates doubt why anyone would ever want to use a multiple system.

(There may be reasons, but it must be shown that they are worth the accompanying price of indeterminacy.) Earlier, I questioned whether one could trust a voting system if its rankings can be so sensitive to which candidates just happen to be presented. Multiple systems not only preserve these serious flaws, they add a host of distrubing new ones! As a personal example, I wonder what the election results for our Departmental Budget Committee or for the MAA really mean?

As a last sample of what can go wrong in elections, we'll visit the world of strategic voting. Responding to the criticism that his method could be manipulated, Borda reportedly asserted that his system was meant only for honest people. This leaves out Chicago and most of the modern world! Strategic behavior exists. Of course, we now know from the important Gibbard [6] and Satterthwaite [7] result that <u>all</u> reasonable (e.g., no dictatorships) election procedures ranking three or more candidates can be manipulated. Therefore, we need a "second best" approach; we need to know what methods [illegible] likelihood of a successful manipulation.

An approach to resolve this question is suggested by our method of proof. Let $g(\underline{p},\underline{W})$ be the $\underline{W}$ tally of an election of the set of N candidates for profile $\underline{p}$. To manipulate the outcome, some voters will assume a different ranking of the candidates; i.e., the election will be tallied with $\underline{p}'$ instead of $\underline{p}$. Here, $\underline{p}' = \underline{p} + \underline{v}$ where $\underline{v}$ indicates the change due to manipulation. If the manipulation is successful, it altered the election ranking of some two candidates. This means that $g(\underline{p},\underline{W})$ is on one side of the tie tally for the two candidates, while $g(\underline{p}+\underline{v},\underline{W})$ is on the other. So, to analyze this problem, look at the hyperplane of profiles in $Si(N!)$ that cause a tie vote between the two candidates. The <u>susceptibility of a voting method</u> is measured by the number of $\underline{p}$'s that are positioned close enough to this hyperplane so that the sincere profile, $\underline{p}$, is on one side while the manipulated profile, $\underline{p}+\underline{v}$, is on the other. Namely, it measures the number of opportunities admiting a successful manipulation.

Notice the similarity of this formulation to a fluid flow problem ($\underline{v}(\underline{p})$) through a higher dimensional membrane (of dimension N!-2). Continuing this analogy, the problem is to minimize the amount of (manipulation) fluid passing through. In this way, the problem can be solved. For instance, with appropriate neutrality assumptions (any profile is equally likely, it is equally likely that any pair is the target of a manipulation, etc.), and for N=3, I've shown that <u>Borda's method minimizes the likelihood of a successful manipulation</u>! [9] (How appropriate; Borda's main research involved fluid flow.) This was surprising to me because it's possible to construct examples where Borda's method is the worst! (So, Borda's method minimizes the number of such examples.) Of course, different assumptions yield different answers. In fact, it turns out that <u>any</u> system can be justified as being the "best" if you impose the appropriate assumptions on how certain voters behave, their predilection to cheat, etc. That's one reason I adopted "neutrality" assumptions.

Although it's obvious, it's worth mentioning that multiple methods provide far more opportunities to manipulate the system than ordinary voting systems. This is because the indifference surface in the fiber space has dimension kN!-2 rather than just N!-2. This extra dimension provides the strategic voter with extra advantages. To illustrate, suppose a profile is too far from the surface for a voter to manipulate the system by using $\underline{W}_{1N}$; however, it could be close enough to successfully manipulate the system with, say, $\underline{W}_{2N}$. In other words, a multiple system not only provides more opportunities to manipulate the system, but it even sanctions the added, new strategies! But, you knew this, and you've used it. When you're a candidate with an election procedure that permits you to vote for several candidates, you probably "truncate" your ballot (to create a multiple system) by voting only for yourself. You know you're trying to manipulate the system to your advantage.

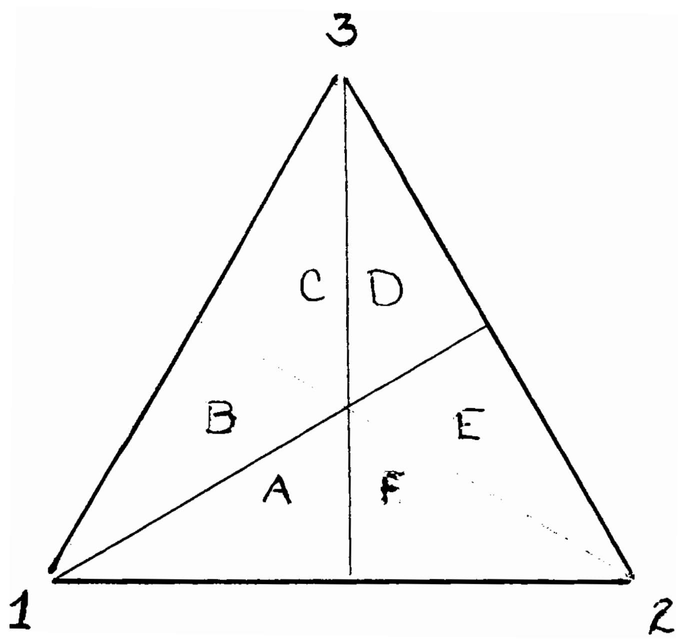
5. Symmetry and the Axiomatic Approach

The principal cause of these paradoxes is that, unknowingly, we're forcing a large domain onto a lower dimensional range; the "squashed" overflow creates the paradoxes. This is only part of the story. Occasionally on this excursion through the perils of the electoral process, I've hinted that part of the difficulty is due to the symmetry of permutation groups. The groups arise by identifying each alternative, (a_j), $j=1,..,N$, with its subscript, so a ranking of the alternatives can be identified with a specified permutation of these symbols. In this way, the ranking $a_2 > a_1 > a_3$ is identified with the permutation $[2,1,3]$ and $a_1 > a_2 > a_3$ with $A = [1,2,3]$. Let $P(1,..,N)$ be the space of all of $N!$ permutations for N alternatives.

The symmetry group, S_N, acts on $P(1,..,N)$ where its orbit of any ranking in $P(1,..,N)$ is all of $P(1,..,N)$. For instance, $[2,1,3] = (1,2)(3)A$. So, by choosing one permutation, say, $A=[1,..,N]$, S_N is identified with $P(1,..,N)$ by identifying B in $P(1,..,N)$ with q in S_N where $qA=B$.

In voting, we're interested not just in $P(1,..,N)$, but also in the rankings of the subsets of alternatives. To describe this, consider $P(1,2,3)$ and $P(1,2)$. A voter defines a ranking in $P(1,2,3)$, and this ranking uniquely determines a ranking in $P(1,2)$; e.g., if $[2,1,3]$ characterizes the voter in $P(1,2,3)$, then the restriction $[2,1]$ is the appropriate ranking in $P(1,2)$. Thus, a voter's ranking is an element of the diagonal D in $P(1,2,3) \times P(1,2)$.

It turns out that <u>many of the negative results from social choice</u> are based on the difficulty in expressing D as the orbit of a group. Clearly, the


3
C
D
B
E
A
F
1
2

group action involves S_3xS_2. But, D isn't an invariant subset of S_3xS_2, so restrictions need to be imposed. To see this, start with the permutation (1,2,3) and consider (1,2,3)[1,2,3] = [2,3,1]. The restriction of the image, [2,1], is the transposition, or flip, of the restriction of [1,2,3], [1,2]. This suggests that associated with (1,2,3) in S_3 is the flip, F, from S_2. It works here because ((1,2,3),F)([1,2,3],[1,2]) = ([2,3,1],[2,1]) maps an element from D to D.

This isn't sufficient. For instance, consider ((1,2,3),F) acting on another element of D, say ((1,2,3),F)([3,1,2],[1,2]) = ([1,2,3],[2,1]). This image is <u>not</u> on the diagonal. To force the image on D, the group action needs to be ((1,2,3),I) where I is the identity. Evidently, the choice of the element from S_2 depends not only on <u>what group action</u> is selected from S_3, but also on <u>what element</u> in P(1,2,3) is acted on. In other words, the group action selected from S_2 is determined by a mapping $s:S_3xP_3 \dashrightarrow S_2$. When certain appropriate assumptions are imposed on s, this defines a new group, $S_3{}^{\sim}S_2$ - the wreath product of the two groups. I'll need the extension to several groups - the iterated wreath product.

To have a better idea of what is going on, consider the geometric representation of the rankings given in Figure 1 where each vertex is identified with one of the alternatives. A point in the simplex represents a binary relationship in the following way: the closer a point is to a vertex, the stronger that alternative is preferred. In this way, the baricentric division of the simplex defines all possible rankings. For instance A = [1,2,3], E = [2,3,1].

The permutation (1,2,3) can be viewed as rotating a ranking 120^o in the counterclockwise direction. For instance, (1,2,3)C = A. This rotation induces a flip in P(1,3) and in P(2,3), but no change in P(1,2). (This can be seen by projecting the ranking onto the rankings on the appropriate edges.) Also,

$(1,2,3)^2C = E$; again, this group action and C define various flips and identity operations on the pairs. So, to understand a change in a voter's rankings, view it as an element from the iterated wreath product $S_3 \sim S_2 \sim S_2 \sim S_2$ operating on the diagonal D_3 in $P_3 = P(1,2,3) \times P(1,2) \times P(1,3) \times P(2,3)$.

This 120° rotation - the simpliest group action available in the plane but not on the line - explains the agenda problem faced by our Chair. Five members of this Department have ranking A, 5 have ranking C, and 5 have ranking E. The tally for any pair is computed by projecting the numbers to the appropriate edge. In this way it's clear what happens - the projection of the symmetry in the plane is a cycle on the edges. With this symmetry, the last listed alternative in an agenda always wins. Alternatively, this cycle corresponds to the number of flip and identity group actions acting on the sets $P(j,k)$ created by the orbit from the wreath products.

More complicated problems can be analyzed in a similar way. This includes the seminal Arrow's Theorem as well as most of the other results from Social Choice. To illustrate this, I'll use my notation to reconstruct a special case of Arrow's Theorem with 3 alternatives and two voters. We are interested in the maps

6. $F: D_3 \times D_3 \dashrightarrow P_3$

where the goal is to characterize the subset of maps that satisfy

7. $\{F: D_3 \times D_3 \text{---}> D_3 \mid F \text{ is onto}\}$.

These are the mappings that admit all possible election rankings where the election ranking of all three alternatives agree with the election rankings of the pairs; i.e., there are no surprises, paradoxes, or cause for Departmental suspicion.

It's easy to find such a mapping; any surjective F from $P(1,2,3) \times P(1,2,3) \text{---}> P(1,2,3)$ defines a ranking of the three alternatives.

Now, let the group's rankings of the pairs be the natural restrictions. But this isn't what we want because it doesn't address the real problem introduced by the beverage example. It just imposes statements that the Departmental members rank milk above wine, etc. To address the real problem, we want the election rankings of any two alternatives to depend only on the voters' relative rankings of these same two alternatives. This means that for all choices of j and k, we require the obvious restriction of F to satisfy

8. $F:P(j,k)\times P(j,k) \dashrightarrow P(j,k)$.

The identity map on one variable satisfies 7 and 8. In our Department, this isn't acceptible because it corresponds to making one of the voters a dictator, but not necessarily a benevolent one. Consequently, our version of Arrow's last assumption is that

9. F cannot be represented by a function of one variable.

Surely, with some imagination, such an F can be found. However

Theorem. The set of mappings satisfying 7,8,9 is empty.

The proof relies on the iterated wreath product of groups. To get a flavor of this, notice that because of 8, to construct such a map you would start with the pairs and let their rankings dictate the full ranking of the triplet. Because of 9, there are situations where each voter can affect the outcome. According to 8, we can assume this occurs with some pair.

With all the symmetry, it doesn't matter which pairs we choose. So, assume the first voter can affect the P(2,3) ranking and the second can affect the P(1,2) ranking. It may be that this is true only when the other voter has a fixed specific ranking of the same two alternatives. For instance, suppose that voters 1 and 2, respectively, need the rankings [1,2] and [2,3] for the above to be true. (By symmetry, everything can be modified for any other choice.) These conditions are satisfied if the first voter's rankings vary between A and B while the second voter's rankings vary between A and F.

According to the above, this means that the election outcome of the pairs varies between [1,2] and [2,1], [2,3] and [3,2] <u>independent</u> of each other. But, the [1,2] and [2,3] outcomes forces the full ranking to be A while the [2,1] and [3,2] rankings forces the ranking to be D. This means that although the two voters <u>never</u> changed their P(1,3) rankings, the P(1,3) election outcome did! This contradicts 8 and proves the theorem.

So, it is the changes in the voters' preferences that are captured by the wreath product of groups. To prove the result, elements from P(1,2,3) and actions from the wreath product are chosen so they define a flip on one pair of the alternatives and the identity on the other two pairs. By 9, this means that the image is characterized by independent flips over two different pairs of alternatives. In turn, this forces the image to flip on the remaining pair, but this cannot occur.

All of the other Social Choice results that I've encountered, including the Gibbard - Satterthwaite theorem and results about restrictions on preferences, are of a similar nature. Certain "independence" conditions, either implicitly or explicitly, require that the rankings over certain subsets of alternatives to be determined by the voters' relative rankings of these same alternatives. This defines the appropriate wreath product of permutation groups. Instead of using the usual combinatoric arguments, these theorems could be prove and extended by showing there doesn't exist a mapping from the product of certain wreath products to another wreath product that satisfies the specified properties. The alternative is that F is a mapping of one variable; i.e, the system is a dictatorship.

Our first, basic theorem about tallying methods is proved with similar techniques. I examined the vector space spanned by $\{f(\underline{p},\underline{W}^N)|p\in Si(N!)\}$ where f gives the actual tally. This space is invariant with respect to the iterated wreath product defined by the $2^N-(N+1)$ permutation groups. So, the proof is based on finding and characterizing all of the invariant subspaces of this group action. In particular, one goal was to find the subspaces of lower dimension. Because of this approach, it is reasonable to expect, and it occurs, that these lower dimensional sets form a stratification much like in singularity theory. (It should, these are the singular sets of the group action.) The second theorem is based on the fact that the set of lowest dimension comes from the Borda Count. Finally, with this approach, the normal bundle for the vector space can be determined. From this, the characterization of any dictionary follows with simple vector analysis arguments.

In summary, it is the symmetry of the Borda Count that accounts for its highly favorable properties. These properties are so strong that I recommend you use the Borda Count for your next election. If you don't, you'll pay for it -- you may need to go to another departmental meeting!

Acknowledgements For the most part, this is a written version of my talk at the MAA meeting for the Michigan section in May, 1986. This research was supported by NSF IST 8415384.

References

1. Arrow, K. J. (1951), Social Choice and Individual Values New York: Wiley.

2. Sen,A. 1986), Social Choice Theory, in Handbook of Mathematical Economics, ed. Arrow, K. J., and M. D. Intriligator, New York:North-Holland.

3. Saari, D. G. (1985), The optimal ranking method is the Borda Count, NU Discussion Paper.

4. Brams, S. and P. Fishburn (1983),

5. Saari, D. G., and J. Van Newenhizen (1985), A case against bullet, approval, and plurality voting, NU Discussion Paper.

6. Gibbard, A. (1973), Manipulation of voting schemes: A general result, Econometrica, 41:587-601.

7. Satterthwaite, M. (1975), Strategy-proofness and Arrow's conditions, Journal of Economic Theory, 10:187-217.

8. Saari, D. G. (1986), Susceptibility to manipulation, NU preprint, Sept. 1986.