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

Optimal Group Decisions

Young, H.P.

IIASA Working Paper

April 1994

Working Paper

Optimal Group Decisions

H. Peyton Young

WP-94-26
April, 1994

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Optimal Group Decisions

H. Peyton Young

WP-94-26
April, 1994

Working Papers are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Abstract

Consider a group of individuals who must rank a set of decisions or choices. If the members of the group disagree, how should their opinions be reconciled into a group ordering? Historically we may discern two ways of answering this question. The "relativist" approach, which is the dominant one in the modern social choice literature, holds that differences of opinion arise largely from differences in preferences or values. Hence the objective should be to strike a fair compromise between differences of opinion. The "rationalist" approach, which was an article of faith among the eighteenth-century founders of voting theory, holds that differences of opinion arise from misperceptions about the true merit of different decisions. For them the goal was to find the ordering that is most likely to be "correct" or "true". These two positions are not incompatible. Under suitable conditions, in fact, they yield the same method, which was first proposed in a rudimentary form by Condorcet. We show that it can be characterized by a slight weakening of the independence of irrelevant alternative condition.

Optimal Group Decisions

Peyton Young

Simple majority rule

"Democracy," said E. B. White, "is the recurrent suspicion that more than half the people are right more than half the time" (White, 1946). The suspicion has been around for a long time. It is, in fact, the central thesis of a remarkable work published in 1785 by the French mathematician and political philosopher, Jean Antoine Nicholas Caritat, Marquis de Condorcet.[1] Condorcet set out to prove that majority rule is not only a *fair* way to make political choices, it is actually the *best* way to do so -- the way most likely to yield "optimal" results. Though this notion may at first sound strange to modern ears, it turns out to be a surprisingly fruitful way of thinking about the design of voting rules.

Condorcet begins with the premise that the object of government is to make decisions that are in the best interest of society. This leads naturally to the question: what voting rules are most likely to yield good outcomes? To analyze this problem he applied the then-novel science of probability theory. Imagine that a group of voters must decide between two alternatives, one of which is objectively *best.* (Whether this is a meaningful assumption will be considered below.) Each individual makes a judgment about this question and registers his opinion. Sometimes, of course, people judge incorrectly. But let us assume -- perhaps too optimistically -- that each voter is *more likely* to make the right choice than the wrong choice in any given situation. Condorcet showed that, if the voters make their choices independently, then the laws of probability imply that the choice with the most votes is the one most likely to be correct. In other words, majority rule is an optimal method of group decision making.

Consider the following example: there are one hundred individuals choosing between two alternatives a and b. Let the outcome be 55 votes for a and 45 votes for b. Assume that each of these individuals is right 60 percent of the time. Now there are two possibilities : either a is really the best choice or b is the best choice. In the first case the observed

[1] *An Essay on the Application of Probability Theory to Plurality Decision Making,* 1785.

voting pattern (55 for a, 45 for b) would occur with probability $(100! / 55! \, 45!) \, .6^{55} \, .4^{45}$, while in the second case it would occur with probability $(100! / 45! \, 55!) \, .6^{45} \, .4^{55}$. As the former is about 58 times more probable than the latter, we conclude that a is more *likely* to be correct than b. This style of reasoning is known in statistics as maximum likelihood estimation. It is somewhat analogous to Bayesian inference with uniform priors, though formally it does not rely on the notion of priors at all.[2]

Why should we buy the idea, though, that there really is such a thing as a *best* choice? Aren't values relative, and isn't the point of voting to determine a compromise between conflicting opinions, all of which are equally entitled to a hearing? This relativistic point of view, which has reigned supreme in social choice theory (indeed in economic theory generally) for most of this century, seems completely at odds with Condorcet's basic premises. Yet Condorcet's point of view is not nearly as absurd as it might seem at first blush. Consider a trial by jury. Let a be the proposition that the defendant is guilty, b that she is innocent. If all twelve members of the jury vote for conviction, and if the probability that each is correct is 0.60, then the probability that the defendant is in fact innocent is less than one in fifty thousand (assuming equal prior probability of guilt or innocence).

This "unanimity rule" is called for if a false conviction is deemed much worse than letting a guilty party go free. But if the objective is simply to reach the *correct* decision with highest probability, then clearly this is not the best one can do. The probability of making the correct choice under unanimity rule is only slightly greater than one-half (about .501). But under simple majority rule the probability of a correct choice is about .665. More generally, it can be shown that, *among all group decision rules on two alternatives, simple majority rule is most likely to yield the correct outcome.*[3] Furthermore, as the size of group becomes larger, the probability that the majority decision is correct approaches unity -- a result first proved by Condorcet.[4]

---

[2] A Bayesian analysis of the situation would proceed as follows. Suppose that a and b are equally likely to be best *a priori*. Let p be the conditional probability that the vote would occur given that a is best, and let q be the conditional probability that it would occur given that b is best. Then the posterior probability that a is best is $p/(p + q) = .983$, while the posterior probability that b is best is $q/(p + q) = .017$. Thus we would choose a rather than b.

[3] For this result we assume that the number of voters is odd, that each makes his judgment independently, and that each chooses correctly with probability p, where $1/2 < p < 1$. See Nitzan and Paroush (1982) and Shapley and Grofman (1984).

[4] For related work see Poisson (1837), Urken and Traflet (1974), Grofman, Owen, and Feld (1983), and Grofman and Feld (1988).

This result reaches far beyond jury trials. It applies, in fact, to any choice problem in which people agree about the objective, but there is disagreement about the best means to achieve that objective. For example, the members of the Federal Reserve Board may all want to maximize the long-run rate of economic growth, but at any given moment they may be uncertain whether lowering or raising interest rates is the best way to achieve this. Similarly, the directors of a corporation may agree that their task is to maximize the firm's long-run profitability, but they may have different views about which candidate for chief executive is most likely to realize this goal. In these situations Condorcet's premises make perfect sense, and simple majority rule is the best way to estimate the optimal choice. Moreover, this is how such decisions are usually made in practice.

Weighted majority rule

Suppose, however, that some voters are known to have better judgment than others. Should we still give all votes equal weight, or should we perhaps ignore all voters *except* those whose judgment has proven to be best in the past? In fact neither answer is correct. To see why, consider a situation in which the five members of a central bank are voting whether to raise interest rates. Assume that each member votes to raise rates if and only if she believes that not increasing rates will cause inflation to exceed the target set by the bank. Assume further that the track record of each member of the committee -- how often she has been right about the effect of interest rates on inflation -- is known from past experience. Perhaps A and B have been right 90 percent of the time, while C, D, and E have been right only 60 percent of the time. If we give all votes equal weight, we ignore the fact that A's and B's opinions convey more accurate information than the others. But if we set aside all opinions except those of A and B, we are losing information from the others that is accurate on the whole. There is value in having more rather than fewer independent judgments. Thus it stands to reason that the opinions of A and B should be given more weight than the others, but not all the weight. It turns out that the optimal decision rule is to use a weighted majority, where the weight on each member's vote is $\log (p_i/(1 - p_i))$ and $1/2 < p_i < 1$ is the probability that individual i makes correct judgments (Nitzan and Paroush, 1982; Shapley and Grofman, 1984). In the present case this means that each of A and B should have about 5.4 times the number of votes that each of D, E, and F has. This amounts to saying that a "majority" is a group that contains both A and B, or A plus at least two of the weak members, or B plus at least two of the weak members.

Extension to three or more alternatives

When we try to extend this reasoning to three or more alternatives, however, matters become more complicated. For, as Condorcet was the first to show, there *is* no direct analog of simple majority rule in this case. The difficulty is that there are situations in which no alternative obtains a simple majority over every other. Consider the following example from Condorcet's *Essay* involving three alternatives and sixty voters. Here majority rule comes up empty-handed because it leads to a voting cycle: a beats b by 33 to 27, b beats c by 42 to 18, and c beats a by 35 to 25.

23	17	2	10	8
a	b	b	c	c
b	c	a	a	b
c	a	c	b	a

Example 1

The problem that Condorcet set himself was to determine the voting rule that is optimal under these circumstances. To make the example concrete, let us suppose that a stands for the policy "hire more police," b stands for the policy "increase prison sentences," and c is the policy "offer training programs for ex-convicts." Each of these policies aims at reducing crime, and we shall suppose that the voters are trying to judge which will be the most effective, i.e., which will reduce the crime rate most per dollar spent. In other words, there really is a correct ranking of the policies, but there are differences of opinion about what that ranking might be.

Assume that, given any two of the policies to compare, each voter has a fixed probability $p > 1/2$ of choosing the right one (i.e., the most effective one). Assume further that each voter's judgment about any pair is independent of the other voters' judgments, and that his judgment about a given pair is independent of his judgment about the other pairs. (One can quibble with the realism of these assumptions but they merely serve to simplify the calculations.)

For each possible ranking of the alternatives, we want to compute the conditional probability that the above voting pattern would occur given that the ranking is correct. Suppose, for instance, that the correct ranking is a b c. Now the probability of the above vote is the product of three terms: the probability that a gets 33 votes over b, the probability that b gets 42 votes over c, and the probability that a gets 25 votes over c. Hence the probability of the vote in Example 1 is proportional to $p^{33}(1 - p)^{27}$ x $p^{42}(1 - p)^{18}$ x $p^{25}(1 - p)^{35} = p^{100} (1 - p)^{80}$. In general, the larger the exponent of p, the higher the likelihood that the given ranking is the correct one. Thus we will have solved the problem once we compute the exponent of p for each of the six possible rankings.

A convenient way to visualize the problem to draw a *vote graph* like that shown in Figure 1.

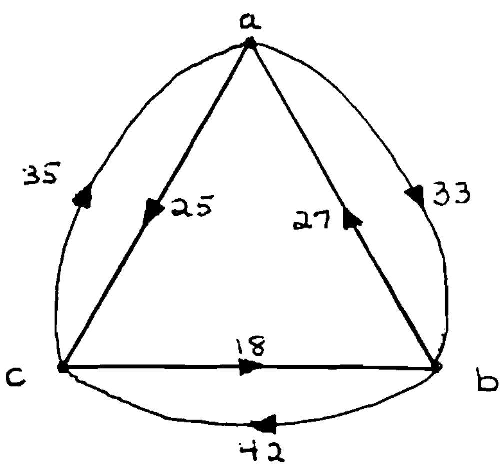

Figure 1.

There is one vertex for each alternative, and between every pair of vertices there are two edges, one in each direction. The *weight* on an edge is the number of votes that the alternative at its base gets over the alternative at its tip. To evaluate the probability of a ranking such as a b c, consider the three pairwise propositions: a is above b, b is above c, and a is above c. These correspond to the three directed edges a $\rightarrow$ b, b $\rightarrow$ c, and a $\rightarrow$ c, which have weights 33, 42, and 25 respectively. Thus the total *support* for the ranking a b c is 33 + 42 + 25 = 100. The support for the other rankings is found in like fashion. Namely, each ranking corresponds to a set of three directed edges that do not form a cycle, and the corresponding exponent on p equals the total weight on these edges. The results are given below, and they show that b c a is most likely to be correct in this case.

a b c	100	b c a	104
a c b	76	c a b	86
b a c	94	c b a	80

In general, whenever there is a voting cycle among three alternatives, the maximum likelihood ranking is obtained by breaking the cycle at its weakest link, that is, by deleting the edge in the cycle that has the smallest majority. (In the above example, the cycle $a \rightarrow b \rightarrow c \rightarrow a$ would be broken at the link $a \rightarrow b$.) When there is no voting cycle, the maximum likelihood ranking is the one that accords with simple majority rule on all pairs of alternatives. This identifies the maximum likelihood ranking when there are exactly three alternatives, and is known as *Condorcet's rule of three.*

The case of more than three alternatives

Unfortunately, when there are more than three alternatives, Condorcet seems to have become confused; at any rate he did not get the correct answer.[5] Nevertheless the solution can be found by a straightforward extension of the previous argument. Given a voting outcome and a ranking R of the alternatives, the conditional probability of observing the vote, given that the true ranking is R, is proportional to $p^{s(R)} (1 - p)^{M - s(R)}$ where $M = nm(m - 1)/2$ and s(R) *is the total number of pairwise votes that higher alternatives get over lower alternatives in R.* We shall call s(R) the *support* for the ranking R. The maximum likelihood rankings are precisely the ones with maximum support. To compute them, it suffices to find *the maximum weight set of edges in the vote graph that does not contain a cycle.* This *maximum likelihood rule* is the solution to Condorcet's problem.[6]

Borda's rule

At this point we need to pick up a second strand in our story that actually begins somewhat earlier. In 1770, some fifteen years before Condorcet published his work on voting theory, his colleague Jean-Charles de Borda read a paper on the design of voting procedures to the French Academy of Sciences. Like Condorcet, Borda was a prominent figure in scientific circles with interests that spanned a wide variety of subjects. Unlike

[5] For a discussion of Condorcet's somewhat obscure argument in this case, see Young (1988) and Crepel (1990).

[6] Young (1986, 1988). The maximum likelihood rule can be formulated as an integer programming problem. Define a variable x_{ij} for each directed edge $i \rightarrow j$ in the vote graph, that is, one variable for each ordered pair of alternatives. Let w_{ij} be the weight on edge $i \rightarrow j$, that is, the number of votes for alternative i over alternative j. A maximum likelihood ranking corresponds to a solution **x** that maximizes $\sum w_{ij}x_{ij}$ subject to $x_{ij} + x_{ji} = 1$, $x_{ij} + x_{jk} + x_{ki} \leq 2$, and all $x_{ij} = 0$ or 1. If the voters have different competences, individual i's vote is weighted by $\log (p_i/(1-p_i))$, where p_i is the probability that i is correct and $1/2 < p_i < 1$.

Condorcet, he had a strong practical bent. In addition to doing important research in mechanics, hydraulics, and optics, he was one of the leading lights in developing the metric system. This put him in the applied faction of the Academy, which was often at loggerheads with purists like Condorcet. This rivalry seems to have spilled over into their work on voting theory, as we shall see later on.

Borda began by observing that, when there are three or more alternatives, the one that achieves the most first-place votes is not necessarily the one that has the most overall support. As an example, consider the following situation. There are three alternatives and twenty-one voters who rank them as follows:

7	7	6	1
b	a	c	a
c	c	b	b
a	b	a	c

Example 2

Under the conventional *plurality method*, a gets 8 votes, b 7, and c 6. But this fails to take into account that all but one of the people who prefer a like c better than b, and everyone who prefers b likes c better than a. One could therefore argue that c is the most natural compromise candidate even though it receives the *fewest* first-place votes. To assess the true strength of the various candidates, said Borda, one must look at their overall standing in the individual rankings. This led him to propose the following rule. Let each voter strictly rank-order the candidates. (For simplicity of exposition we assume no indifference.) In each voter's list, assign a score of 0 to the alternative ranked last, a score of 1 to the alternative ranked next-to-last, a score of 2 to the alternative next above that, and so forth. The *Borda score* of an alternative is its total score summed over all voter lists, and *Borda's rule* is to rank the alternatives from highest to lowest Borda score. In the above example the scores are 26 for c, 21 for b, and 16 for a. Thus, according to Borda's rule, the proper ordering of the alternatives is c b a, which is exactly the *opposite* of the one implied by the number of first-place votes.

Although it may not be obvious at first, Borda's rule (like Condorcet's) actually depends only on the *pairwise votes* between the various alternatives. The simplest way to see this is to observe that the Borda score of an alternative within a particular voter's list is just the number of alternatives ranked below it. It follows that the total Borda score of an alternative x is the total number of times that x beats other alternatives, summed over all of the voter lists.

A convenient way of computing the Borda solution is to construct a matrix indicating the pairwise votes. For each ordered pair of alternatives a, b, let v_{ab} denote the total number of times that a is ranked above b in the voter lists. In other words, v_{ab} is the number of votes that a would get over b in a pairwise contest between the two, assuming that everyone votes according to his preferences. (This is the weight on the edge directed from a to b in the vote graph.) The m x m matrix (v_{ab}) is called the *vote matrix*. It follows from the above discussion that the Borda score of each alternative x is just the sum of the entries in x's row. In terms of the vote graph, it is the sum of the weights on all edges directed away from the vertex corresponding to x. The vote matrix for Example 2 is shown below, together with the row sums.

	a	b	c	Row sum
a	--	8	8	16
b	13	--	8	21
c	13	13	--	26

Condorcet's critique of Borda: independence of irrelevant alternatives

Borda's paper was not published until 1784, one year before Condorcet's treatise on voting appeared. Condorcet took strong exception to Borda's method on the merits; moreover he added a certain amount of personal venom to the attack. (Throughout the *Essay* Condorcet refers sarcastically to the "method of a famous mathematician" but fails to mention him by name. The implication is that the method is not worthy of a mathematician.) [7]

[7] Condorcet's contemptuous view of Borda is borne out in his private correspondence. For example, in a letter to Turgot he wrote: "[M. Malesherbes] makes a great case for Borda, not because of his memoirs, some of which suggest talent (although nothing will ever come of them, and no one has ever spoken of them or ever will) but because he is what one calls a good academician, that is to say, because he speaks in meetings of the Academy and asks for nothing better than to waste his time doing prospectuses, examining machines, etc., and above all because, feeling eclipsed by other mathematicians, he, like d'Arcy, has abandoned mathematics for petty physics." (Henry, 1883).

To illustrate his objections to Borda's method, Condorcet introduced the following example. There are three candidates named Peter, Paul, and Jack, and eighty-one voters with the following preferences:

30	1	29	10	10	1
Peter	Peter	Paul	Paul	Jack	Jack
Paul	Jack	Peter	Jack	Peter	Paul
Jack	Paul	Jack	Peter	Paul	Peter

Example 3

According to Borda's rule the proper ordering is Paul, Peter, Jack. But this is absurd, said Condorcet, because Peter obtains a simple majority over *both* Paul *and* Jack. Surely this means that Peter is stronger than Paul. More generally, Condorcet formulated the *majority principle*, which states that *if there exists an alternative that obtains a simple majority over every other alternative, then it should be ranked first.* Such an alternative is known as a *Condorcet alternative* or a *majority alternative.* Condorcet went on to examine why Borda's rule gives the "wrong" result in this case.

> "[H]ow is it that Paul is not the clear winner when the only difference between himself and Peter is that Peter got 31 first places and 39 second, while Paul got 39 first and 31 second? Well, out of the 39 voters who put Peter second, 10 preferred him to Paul, whereas only one of the 31 voters who put Paul second preferred him to Peter. The points method [of Borda] confuses votes comparing Peter and Paul with those comparing either Peter or Paul to Jack and uses them to judge the relative merits of Peter and Paul. *As long as it relies on irrelevant factors to form its judgments, it is bound to lead to error*, and that is the real reason why this method is defective for a great many voting patterns, regardless of the particular values assigned to each place. The conventional method [plurality] is flawed because it ignores elements which should be taken into account and the new one [Borda's] because it takes into account elements which should be ignored" (italics added). [8]

[8] This passage is from a later paper of Condorcet's entitled "On the Constitution and the Functions of Provincial Assemblies" (Condorcet, 1788). The translation is by Sommerlad and McLean (1989), who were the first to call attention to its connection with independence of irrelevant alternatives.

In other words, Condorcet is saying that the comparison between Peter and Paul should depend only on the relative ordering of these two candidates in the voters' lists, not on their relation to other candidates. Here is a clear forerunner of Arrow's principle of the independence of irrelevant alternatives!

Condorcet noted the remarkable fact that *any* scoring system leads to the same outcome as Borda's rule in this example, and is therefore subject to the same criticism. In general, a *scoring method* is defined by a sequence of real numbers $s_1 > s_2 > \ldots > s_m$, one for each alternative. Given an individual's ranking of the alternatives, assign a score of s_1 to the alternative that occupies first position, a score of s_2 to the alternative in second position, and so forth. The total score of each alternative is the sum of its scores over all voter lists, and the alternatives are ordered according to their total scores. Borda's rule corresponds to the scoring system $s_i = m - i$; in fact, it is equivalent to any scoring system in which the successive differences $s_i - s_{i+1}$ are equal and positive.

Consider now any scoring system for three alternatives with descending scores $s_1 > s_2 > s_3$. In example 3, the score for Peter is $31s_1 + 39s_2 + 11s_3$ whereas the score for Paul is $39s_1 + 31s_2 + 11s_3$. Therefore, Paul obtains a higher score than Peter, even though Peter is the majority candidate. From this we conclude that *any* scoring system violates the majority principle. Moreover, it shows that any scoring system yields outcomes that are based on "irrelevant factors."

Local independence of irrelevant alternatives

Condorcet's broadside against Borda is fine as far as it goes. But exactly how far does it go? We know from Arrow's theorem that independence of irrelevant alternatives is violated by *every* reasonable decision rule when there are more than two alternatives. Why then should we believe that Condorcet's approach is any better than Borda's? In this section we shall show that it is, in fact, considerably better.

To motivate the discussion, let us first consider why a condition like independence of irrelevant alternatives is worth having at all. Essentially it says that the way a given group of alternatives is ordered should depend only on opinions about those alternatives.[9] There are at least two reasons why this is desirable from a practical standpoint. First, if it

[9] I neglect various fine distinctions in the way that one formulates the independence condition.

does not hold, then it is possible to manipulate the outcome by introducing extraneous alternatives (Gibbard, 1973; Satterthwaite, 1975). Second, independence allows the electorate to make sensible decisions within a restricted range of choices without worrying about the universe of all possible choices. It is desirable to know, for example, that the relative ranking of candidates for political office would not be changed if purely hypothetical candidates were included on the ballot.

While Arrow's theorem shows that independence cannot be fully realized by any democratic rule, we shall show that it can be realized to a significant extent. Consider Example 3 again. The real contest here is between Peter and Paul; Jack is a distinctly weaker alternative. We could argue that Jack ought to be "irrelevant" to the choice between Peter and Paul because Jack is inferior to both. Moreover, under Condorcet's rule of three, this is actually the case.

The key point here is that Peter and Paul occur *together* in the consensus ranking; they are not separated by other alternatives. More generally, an *interval* of an ordering is any subset of alternatives that occurs in succession in that ordering. Suppose we insist that, whenever a set of alternatives forms an interval of the consensus preference ordering, then independence of irrelevant alternatives applies -- that is, the ordering within the interval remains fixed when alternatives outside the interval are ignored. In particular, the ordering of alternatives toward the top of the list is unaffected by the removal of those at the bottom. Similarly, the ranking of the alternatives toward the bottom of the list is unaffected by the removal of those at the top, and so forth. We shall say that such a ranking rule satisfies *local independence of irrelevant alternatives* (LIIA). It is a remarkable fact that the maximum likelihood method satisfies LIIA.[10] Moreover, as we shall argue in a later section, it is the only reasonable ranking rule that does so.

To further illustrate this idea, consider the following vote matrix involving six alternatives and one hundred voters.

[10] See Young (1988), where the LIIA condition was called *local stability*.

	a	b	c	d	e	f	Sum
a	--	51	54	58	60	62	285
b	49	--	68	56	52	58	283
c	46	32	--	70	66	75	289
d	42	44	30	--	41	64	221
e	40	48	34	59	--	34	215
f	38	42	25	36	66	--	207

Example 4

Here we may think of a, b, and c as being the real choices under discussion, while d, e, and f are red herrings that have been dragged in by political strategists to attempt to manipulate the outcome. (Note that each of a, b, c has a majority over each of d, e, f, so the latter three are weaker than the former.) Moreover, this attempt will succeed if Borda's rule is used. The row sums determine the Borda ordering c a b d e f. Now suppose that the three red herrings had not been introduced into the debate. Then the vote matrix would be the one enclosed by the dashed lines. The Borda scores for this three-alternative situation are 105 for a, 117 for b, and 78 for c. Hence, in the absence of d, e, f, that top three alternatives would be ordered b a c. But this is exactly the *reverse* of how they are ordered when all six alternatives are considered together. This example shows why Borda's rule is highly susceptible to manipulative practices, just as Condorcet alleged.

Now consider the maximum likelihood (ML) solution to this problem. We begin by observing an important property of this rule: *if some alternative has a simple majority over every other, then the ML rule must rank it first.* The reason is simple. Suppose that x (the majority alternative) were not ranked first. Then it must be ranked immediately below some other alternative y. By assumption, x defeats y by a simple majority. Therefore, if we switch the positions of x and y, we obtain a new ranking that is supported by more pairwise votes. But then the new ranking is *more likely* than the original ranking, which is a contradiction.

This simple fact can be used to deduce the ML solution to the above problem almost immediately. Since a is the majority alternative, it must be ranked first. Among the remaining alternatives, b obtains a simple majority over c, d, e, and f. Hence a similar argument shows that b comes next. Then comes c by the same argument. As for d, e, and f, they will be ordered relative to one another *as if they were the only three alternatives.* (Since they form an interval of the consensus ordering, LIIA applies). It is easy to see that the ML solution for these three alternatives is d f e. Putting all of this together, we conclude that the ML solution to the whole problem is a b c d f e. This example shows that the ML solution is often quite easy to calculate even when there is more than a handful of alternatives.

At this point we can also see more clearly why the method satisfies LIIA, that is, why *any interval of alternatives is ranked as it would be in the absence of the others.* The reason is this: if it were not, then one could shuffle the alternatives *within the interval* and obtain an ordering that is supported by a larger number of pairwise votes, hence an ordering that has greater likelihood. This contradiction shows that the maximum likelihood rule satisfies local independence of irrelevant alternatives. In particular, it cannot be *manipulated from below* by introducing inferior alternatives, nor can it be *manipulated from above* by introducing utopian (i.e., attractive but infeasible) alternatives.

The maximum likelihood method as a form of compromise

So far we have proceeded on the premise that there really is a best ordering to be estimated. Moreover we have argued that this is often the right way to think about group choice problems. But it is not always the right way to think about them. There are many situations in which differences of opinion do not arise from erroneous judgments, but from differences in values. In this case it seems better to adopt the view that group choice is an exercise in defining consensus, that is, in finding a compromise between conflicting opinions. Arrow's axiomatic approach is one way of analyzing this issue, and we shall pick up this scent again in the next section. First, however, I want to draw attention to another interesting approach along these lines that was pioneered by John Kemeny (1959).

Kemeny viewed the voters' opinions as data, and asked for the ordering that best *represents* or *averages* the data. For the notion of average to make sense, of course, we must have some way of measuring how far apart one ranking is from another, that is, we need a metric defined on the set of rankings. Kemeny proposed the following natural metric: *the distance between two rankings R and R' is the number of pairs of alternatives* *on which they differ.* Thus if R' is obtained from R by interchanging two adjacent alternatives, then d(R, R') = 1. If R' is obtained from R by reversing the order of all alternatives, then d(R, R') = m(m - 1)/2, and so forth.[11]

Suppose now that each member of a group of n voters submits a ranking of the alternatives. Given these n data points, what is the best definition of a compromise ordering? A statistician would say there are two obvious answers: the mean and the median. The *mean* is the ranking (or rankings) that minimize the sum of squares of distances from the n given rankings. The *median* is the ranking (or rankings) that minimize the sum of distances from the n given rankings. It is not difficult to show that the median is equivalent to the maximum likelihood method, whereas the mean yields a quite different scheme.

Kemeny left open the question of whether the mean or the median was to be preferred. There can be little doubt, however, that the median is the better choice under the circumstances. To see why, consider the following example with 41 voters and three alternatives.

21	5	4	11
a	b	c	c
b	c	a	b
c	a	b	a

Example 4

Alternative a has an absolute majority of first-place votes, so *a fortiori* it is the majority alternative. Indeed, the ranking a b c is supported by a majority, so it has maximum likelihood, that is, it is the median ranking. (It may be checked that Borda's rule yields the same result.) A simple calculation shows, however, that the *mean* ranking is b a c. This seems to be a less credible conception of consensus than does a b c. The problem with the mean is that it places a lot of weight on extreme observations. In the present case, the voters who announce the ordering c b a shift the outcome in favor of b, not because they are especially attached to b, but because their top candidate is c, which is at

[11] Actually, Kemeny defined the distance between two rankings to be twice the number of pairs on which they differ. He also defined the distance between orderings with indifference. For simplicity of exposition we ignore this case.

odds with the views of most of the other voters. In other words, their opinion about b versus a is heavily weighted because their opinions about something else (namely c) differs from the opinion of the majority. This does not seem very sensible. We conclude that, if the object is to find a *compromise* between the various rankings reported by the voters, then the median is, in a statistical sense, the most appropriate solution. This reinforces the argument for the maximum likelihood rule, but from a different (and more modern) point of view.

An axiomatic justification

Pursuing this point of view a bit further, one might wonder whether the ML rule can be justified from purely axiomatic principles. We shall show that it can. Indeed, it is the unique ranking rule that satisfies three standard axioms in the social choice literature plus local independence of irrelevant alternatives.

Define a *ranking rule* to be a function that associates one or more consensus rankings with every set of rankings reported by a group of individuals on a finite set of alternatives. The rule is *anonymous* if it treats all voters alike. It is *neutral* if it treats all alternatives alike. It is *Pareto* if, whenever everyone ranks one alternative above another, then so does the consensus ranking. Finally, a rule *satisfies reinforcement* if, whenever two distinct groups of voters reach the same consensus ordering under separate votes, this ordering is also the consensus for the two groups merged together. For example, if the House of Representatives orders three choices a b c, and the Senate also orders these choices a b c (using the same voting rule), then a b c is the outcome when the rule is applied to both houses together and the votes remain as before.[12] (In practice almost all rules have this property.) It may then be shown that *the maximum likelihood rule is the unique ranking rule that is anonymous, neutral, Pareto, satisfies reinforcement, and local independence of irrelevant alternatives.* [13]

[12] This idea was introduced by Young and Levenglick (1978). In the case of ties, reinforcement states that the rankings chosen by *both* groups separately (if any such exist) are precisely the rankings that result when the votes of the two groups are pooled. A variation of the concept characterizes scoring methods (Smith, 1973; Young, 1974, 1975).

[13] This follows from Young and Levenglick (1978), Theorem 3.

Conclusion

We conclude that the maximum likelihood method for ranking alternatives can be justified from several different points of view. On the one hand it is arguably the best method if we think of voting as a collective quest for truth, that is, as a way of estimating what decisions are most likely to be "correct" or most likely to meet a given objective. This is quite a common situation, especially when the decision is being taken by a group of experts. But it also applies to many forms of political decision making -- what bill is most likely to reduce crime, what foreign policy will minimize the prospect of war, and so forth.

On the other hand, there are surely situations where it is more natural to think of voting as a way of compromising between conflicting values. Here again the ML rule makes sense, because it represents the median opinion. Alternatively, one might want a method that is resistant to strategic manipulation. As we know from the Gibbard- Satterthwaite theorem, almost no method has this property, which is intimately connected with independence of irrelevant alternatives. It is possible, however, to design ranking rules that are immune to manipulation from above by utopian alternatives, and from below by inferior alternatives. Moreover, the ML rule is the only rule that satisfies this local independence of irrelevant alternatives property plus other standard conditions.

The one remaining question is whether the maximum likelihood method is really practical. Compared with more traditional methods like plurality voting, single transferable vote, or even Borda's rule, it is more complicated to calculate. It even seems to have eluded Condorcet's considerable computational skills. Given our present understanding of the problem and modern computing capabilities, however, the issue is largely moot. To find the maximum likelihood solution for six or fewer alternatives, for example, is a near triviality. Even for much larger numbers of alternatives, the method can be implemented in about the same amount of time that it takes people to cast their votes. The more important issue is whether the method is intuitively easy to grasp, and whether it improves on methods currently in use. On both of these counts I think that the answer is affirmative, and I predict that the time will come when it is considered a standard tool for political and group decision making.

References

Arrow, K.J. 1951: *Social Choice and Individual Values.* New York: Wiley.

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

Borda, J.-C. de 1784: Memoire sur les elections au scrutin. *Histoire de l'Academie Royale des Sciences*. Paris.

Condorcet, Marquis de. 1785: *Essai sur l'Application de l'Analyse a la Probabilite des Decisions rendues a la Probabilite des Voix.* Paris: L'Imprimerie Royale. In A. Condorcet O'Connor and M. F. Arago (eds.) *Oeuvres de Condorcet.* Paris. 1847-9.

Condorcet, Marquis de 1788: On the constitution and the functions of provincial assemblies. In A. Condorcet O'Connor and M. F. Arago (eds.) *Oeuvres de Condorcet.* Paris. 1847-9.

Crepel, P. 1990: Le dernier mot de Condorcet sur les elections. Paris: Centre d'Analyse et de Mathematique Sociales.

Gibbard, A. 1973: Manipulation of voting schemes: a general result. *Econometrica*, 41, 587-601.

Grofman, B. and Feld, S. 1988: Rousseau's general will: a Condorcetian perspective. *American Political Science Review*, 82, 567-76.

Grofman, B., Owen, G., and Feld, S. 1983: Thirteen theorems in search of the truth. *Theory and Decision,* 15, 261-78.

Henry, C. (ed.) 1883: *Correspondance Inedite de Condorcet et de Turgot 1770-1779.* Paris: Charavay Freres.

Kemeny, J. 1959: Mathematics without numbers. *Daedalus*, 88, 571-91.

Nitzan, S., and Paroush, J. 1982: Optimal decision rules in uncertain dichotomous situations. *International Economic Review,* 23, 289-97.

Poisson, S.-D. 1837: *Recherches sur la Probabilite des Jugements en Matiere Criminale at en Matiere Civile, Precedees des Regles Generale du Calcul des Probabilites.* Paris: Bachelier.

Satterthwaite, M. 1975: Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10, 187-217.

Shapley, L.S. and Grofman, B. 1984: Optimizing group judgmental accuracy in the presence of interdependencies. *Public Choice*, 43, 329-43.

Smith, J. 1973: Aggregation of preferences with variable electorate. *Econometrica*, 41, 1027-41.

Sommerlad, F., and McLean, I. 1989. *The Political Theory of Condorcet.* Social Studies Faculty Centre Working Paper 1/89. Oxford University.

Urken, A. B. and Traflet, S. Optimal Jury Design. *Jurimetrics*, 24, 218-35.

White, E. B. 1946: The Wild Flag. *The New Yorker.*

Young, H. P. 1974: An axiomatization of Borda's rule. *Journal of Economic Theory*, 9, 43-52.

Young, H. P. 1975: Social choice scoring functions. *SIAM Journal on Applied Mathematics*, 28, 824-38.

Young, H.P. 1986: Optimal ranking and choice from pairwise comparisons. In B. Grofman and G. Owen (eds.), *Information Pooling and Group Decision Making*, Greenwich, Conn.: JAI Press.

Young, H. P. 1988: Condorcet's theory of voting. *American Political Science Review*, 82, 1231-44.

Young, H. P. and Levenglick, A. 1978: A consistent extension of Condorcet's election principle. *SIAM Journal on Applied Mathematics,* 35, 285-300.