Center for Mathematical Studies in Economics and Management Science
Northwestern University, Evanston, Illinois 60208

Discussion Paper No. 928

PROPORTIONAL REPRESENTATION, APPROVAL VOTING, AND COALITIONALLY STRAIGHTFORWARD ELECTIONS

by

Roger B. Myerson

March 1991

Abstract. This paper considers basic constitutional questions about how to elect a legislature. Electoral systems that require blocks of voters to coordinate their votes create a need for pre-election leadership and raise barriers to entry against new parties. Such barriers to entry can rigidify the political system and decrease the incentives for established political leaders to serve the public honestly and effectively. So we consider an axiom of coalitional straightforwardness, which asserts that an electoral system should minimize the need for pre-election coordination of voters who share simple dichotomous preferences. Axioms of nondivisiveness, neutrality of party labels, responsiveness, and homogeneity (or coalitional autonomy) are also formulated. It is shown that only two kinds of electoral systems satisfy these axioms: winner-take-all approval-voting systems (AV), and single-divisible-vote proportional-representation systems (PR). Thus, AV and PR are seen to be uniquely compatible in terms of the incentives for party structure that they create, although AV and PR may differ in their incentives for party positioning in issue space. Possible use of AV and PR together in a bicameral legislature is discussed.

PROPORTIONAL REPRESENTATION, APPROVAL VOTING,

AND COALITIONALLY STRAIGHTFORWARD ELECTIONS

1. Introduction.

The ability of organizations to exploit the broader public is limited by potential competition from other similar organizations. For these purposes, however, the absolute number of such competing organizations that exist at any point in time is not necessarily the definitive structural parameter, because the leaders of a small number of organizations could reach a collusive agreement to jointly exploit the public and then divide the profits. Instead, the critical parameter may be the ease with which new competing organizations could be created if the leaders of existing organizations did collude to exploit the public. Structural variables that increase the costs of creating new competing organizations are called *barriers to entry*. So lowering barriers to entry may be the most effective way of preventing organizations from exploiting the public. In economic theory, where the organizations in question are oligopolistic firms, this significance of barriers to entry has been the subject of much analysis (see Baumol, Panzar, and Willig [1986]). However, the analysis of barriers to entry may be equally important in political science, in the study of parties and other political organizations.

The need of voters in a block or interest group to coordinate their votes on election day can create barriers to entry against new political parties. For example, suppose that 60% of the electorate are leftist voters, 40% are rightist voters, and there are two leftist candidates and one rightist candidate running for an office. Under plurality voting, if the leftist voters divide their votes evenly among the two leftist candidates, then the rightist candidate will win. So, before the election, the leftist voters need some leadership to coordinate them behind one leftist candidate. Thus, the outcome of the election may be determined by those whom the voters accepted as leaders or political arbiters before the election. This power of preelection leaders can create a barrier to entry against new political organizations that attempt to challenge the old leaders. Thus, barriers to entry can be created by voters' incentive to coordinate their votes.

To put it another way, a bandwagon effect (as in Simon [1954]) may be created by voters' fear of wasting votes on parties whose expected support from other voters is small. This bandwagon effect is a barrier to entry against minor parties that do not have a bandwagon rolling for them.

However, bandwagon effects and other incentives for voters to coordinate may be different under different electoral systems. In this paper, we search for procedures for electing a legislature that can minimize such barriers to entry. We consider electoral systems that satisfy a <u>coalitional straightforwardness</u> property which asserts that, at least among a class of voters who have a particularly simple kind of preference over electoral outcomes, there should be no need for coordination and pre-election leadership. A formal definition of this property is developed in Section 3.

Some electoral systems can create underdog effects which effectively reward blocks that divide their support among many small parties. Such rewards are <u>barriers to consolidation</u> for parties. (For an extreme example, consider negative plurality voting, in which each voter names one candidate, and the candidate who is named by the <u>smallest</u> number of voters wins. In this case, if 60% of the voters are leftists and 40% are rightists, and one leftist candidate is running against two rightist candidates, then a rightist candidate

is sure to win, because the negative leftist votes will be split.)

However, there is an integrative function of politics that would be badly served if a block of voters would be penalized for consolidating its support behind one party. Thus, we should seek electoral systems that are also _nondivisive_, in the sense that they create no barriers to consolidation for parties. A nondivisiveness axiom is formulated in Section 4.

In Section 5, we formulate several other properties of legislative electoral systems that may be desirable: _neutrality with respect to party labels_, _responsiveness_, and _homogeneity_ (or _coalitional autonomy_). The main result of this paper, presented in Section 6, is that only two classes of electoral systems satisfy the axiomatic properties from Sections 3, 4, and 5: single-vote proportional-representation systems, and approval-vote winner-take-all systems. Thus, proportional representation and approval voting are seen to be uniquely compatible in terms of the incentives for party structure that they create. Other factors influencing barriers to entry are discussed in Section 7, and possible use of proportional representation and approval voting together in a bicameral legislature is discussed in Section 8. The proof of the main theorem is given in Section 9.

Riker [1982a] has distinguished two kinds of criteria for evaluating electoral systems. Criteria of the first kind, which Riker calls _populist_, evaluate electoral systems by the way that they determine governmental policy positions as a function of voters' preferences. Criteria of the second kind, which Riker calls _liberal_, evaluate electoral systems by the extent to which they restrain the power of government leaders, by forcing them to compete for votes. When we seek electoral systems with low barriers to entry, we are applying a criterion that is purely liberal, in this sense. Our concern is that, whatever governmental policy position might ultimately result from the political process, the electoral system should not deter several independent political parties from advocating essentially the same policy position and competing with each other for public offices and power.

Although proportional representation and approval voting share important properties from this liberal viewpoint, they may seem quite different by populist criteria. Analysis suggests that approval voting would encourage parties to advocate centrist positions (see Brams and Fishburn [1983], Cox [1985], and Myerson and Weber [1988]), whereas proportional representation can encourage parties to take widely divergent positions (see Austin-Smith and Banks [1988]).

2. A general model of legislative elections.

When electoral systems are compared by empirical analysis, it is only possible to compare systems that have actually been used. But we may naturally ask whether some other electoral system that has never been used might satisfy all of our criteria (whatever they may be) better than the systems that have been used in the past. To try to address this question, this paper follows an axiomatic approach to the design of electoral systems. We begin, in this section, by considering a very general class of systems for electing legislative representatives. Then, in Sections 3 - 6, we develop a list of formal properties that such electoral systems might be asked to satisfy, and we characterize the electoral systems that satisfy these properties.

Let us consider an election in which voters in a given district elect legislative representatives. The "district" here could be an entire nation, or a smaller geographical region. We assume that the district has a fixed number of legislative seats, which will be allocated among various competing parties in a way that depends on the results of the voting. We allow the possibility that "seats" may be divisible, and so a party could be allocated any fraction of the district's seats in the legislature.

Let K denote the set of parties that are competing for seats. We assume that K is a nonempty finite set including at least three parties, so

$$|K| \geq 3.$$

Let L denote the set of all proper subsets of the set of parties. That is,

$$L = \{S|\ S \subset K,\ S \neq \emptyset,\ S \neq K\}.$$

We may refer to any set S in L as a (possible) coalition of parties.

Electoral systems like plurality voting and proportional representation often stipulate that each voter must write on his ballot the name of one party, to which he is giving his vote. However, there exist strategically equivalent versions of plurality voting and proportional representation in which each voter has a single divisible vote, which he can divide equally among any set of parties. That is, the set permissible ballots in proportional representation or plurality voting can include all of L if, for each set S in L, we interpet a ballot S to mean the same thing as giving $1/|S|$ votes to each of the parties in the set S. (Here, for any set S of parties, we let $|S|$ denote the number of parties in S.) Thus, in such an electoral system, if all the voters in some block or group submitted the ballot "{1,2}", the effect would be the same as if half of the block voted for party 1 and the other half voted for party 2; so adding the possibility of naming both parties 1 and 2 on the ballot does not increase the range of strategic options that are available to the block. (See Brams and Fishburn [1990] for a discussion of other ways to use coalitional voting in proportional representation systems.)

In general, we let V denote the set of <u>permissible ballots</u> that a individual voter can submit in the election, as specified by the rules of the electoral system. In this paper, we assume that V is a nonempty finite set that includes all of the possible coalitions in L, and so

$$V \supseteq L.$$

That is, we assume that the rules of the election allow the possibility that a voter may submit a ballot which simply lists a set of parties. By itself, the assumption that V includes L can have no restrictive content, because we have not yet made any assumptions about how specific ballots are to be interpreted. (For instance, a ballot that lists two or more parties could be treated like an abstention in some electoral systems. Notice also that V may include any number of possible ballots that are not in the set L.) We are making this assumption here only to simplify the notation in Sections 3 through 5, where restrictive assumptions will be introduced.

For any finite set A, we let $\Delta(A)$ denote the set of all distributions on A. Thus, the set of possible distributions of ballots in the election is $\Delta(V)$, where

$$\Delta(V) = \{\mu \in \mathfrak{R}^V \mid \Sigma_{v\in V}\, \mu(v) = 1,\ \mu(w) \geq 0,\ \forall w \in V\}.$$

The set of possible distributions of seats to the various parties is $\Delta(K)$, where

$$\Delta(K) = \{r \in \mathfrak{R}^K \mid \Sigma_{i\in K}\, r_i = 1,\ r_j \geq 0,\ \forall j \in K\}.$$

(Here, for any finite set A, we let $\mathfrak{R}^A$ denote the set of functions from A into the real numbers $\mathfrak{R}$. Mathematically, $\mathfrak{R}^A$ is a vector space with $|A|$ dimensions.) For any permissible ballot w in V, we let [w] denote the ballot distribution in which everyone is submitting the ballot w. That is,

$$[w] \in \Delta(V), \quad [w](v) = 0 \quad \text{if} \quad v \neq w, \quad \text{and} \quad [w](w) = 1.$$

For any distribution μ in $\Delta(V)$, we may then write

$$\mu = \Sigma_{w \in V} \; \mu(w)[w].$$

We assume that the outcome of the election will depend only on the relative numbers of voters who submit each of the possible ballots. Thus, the rule for determining the outcome of the election can be represented by a function from $\Delta(V)$, the set of ballot distributions, into $\Delta(K)$, the set of parties' seat distributions. We may denote this <u>electoral outcome function</u> by $F:\Delta(V) \to \Delta(K)$. So suppose that μ denotes the distribution of ballots submitted in the election, in the sense that, for each w in V, $\mu(w)$ is the fraction of the electorate that submitted the ballot w. Then $F(\mu) = (F_i(\mu))_{i \in K}$ denotes the distribution of legislative seats to the parties, according to the rules of the electoral system (V,F). That is, for each party i in K, $F_i(\mu)$ denotes the fraction of the district's seats that will be allocated to party i, when μ is the distribution of ballots submitted by the voters. Notice that

$$\Sigma_{i \in K} \; F_i(\mu) = 1 \quad \text{and} \quad F_j(\mu) \geq 0, \quad \forall j \in K, \quad \forall \mu \in \Delta(V),$$

because a party cannot get a negative allocation of seats, and the parties in K will divide among themselves 100% of the available seats. The set of permissible ballots V and the outcome function $F:\Delta(V) \to \Delta(K)$ together completely characterize the electoral system.

(This assumption that the relative distribution of seats depends only on the relative distribution of ballots can itself be derived from two more basic assumptions. First, we may suppose that voters submit their ballots <u>anonymously</u>. This first assumption implies that the distribution of seats can only depend on the numbers of voters who submit each kind of ballot. Second, we may suppose that the relative distribution of seats would not change if this district were enlarged by merging it with other districts in which the absolute numbers of voters submitting each ballot were identical with the original district. This second assumption implies that the outcome can only depend on the relative distribution of ballots, and not on the absolute numbers in the electorate. See also Young [1974] and Smith [1973].)

For example, consider the version of plurality voting in which each voter gets a single vote that he can divide equally among any set of parties. Let $N(\mu)$ denote the set of parties that get the most votes when μ is the ballot distribution; that is,

$$N(\mu) = \{i \in K \mid \Sigma_{S \supseteq \{i\}} \mu(S)/|S| = \max_{j \in K} \Sigma_{S \supseteq \{j\}} \mu(S)/|S|\},$$

Suppose that seats are divided among the winning parties equally, in case of a tie. Then the distribution of seats depends on the distribution of ballots according to the function F where

$$\begin{aligned} F_i(\mu) &= 1/|N(\mu)| \quad \text{if} \quad i \in N(\mu), \\ &= 0 \qquad\qquad\ \ \text{if} \quad i \notin N(\mu). \end{aligned}$$

We may say that an electoral system (V,F) is a <u>proportional representation</u> (<u>PR</u>) system iff, when each voter names a single party on his ballot, the legislative seats are allocated to each party in proportion to the number of voters who named the party. That is, (V,F) is a proportional representation system iff,

(1) $\quad F_i(\Sigma_{j \in K} r_j[\{j\}]) = r_i, \quad \forall i \in K, \quad \forall r \in \Delta(K).$

(Given that r is a distribution on the set of parties, $\Sigma_{j \in K} r_j[\{j\}]$ denotes the ballot distribution in which each voter writes the name of only one party on his ballot and, for each party j, r_j is the fraction of the voters who have named party j.)

There are a variety of ways that proportional representation systems can be extended to the case where voters can also submit ballots that name two or more parties. As we have seen, the simplest way is to suppose that a voter who lists any set S of parties is giving $1/|S|$ votes to each of the parties in the set S. Then the rule of allocating seats to parties in proportion to their vote totals gives us the outcome function F such that

(2) $$F_i(\mu) = \Sigma_{S \supseteq \{i\}} \mu(S)/|S|, \quad \forall i \in K, \quad \forall \mu \in \Delta(L).$$

In approval voting, each voter can give a whole approval vote to each of as many parties as he wants, and the district's legislative seats are all allocated to the party or parties that get the most approval votes. For any ballot distribution μ in $\Delta(L)$, let $M(\mu)$ denote the set of the parties that get the most approval votes when μ is the distribution of ballots; that is,

$$M(\mu) = \{i \in K | \ \Sigma_{S \supseteq \{i\}} \mu(S) = \max_{h \in K} \Sigma_{S \supseteq \{h\}} \mu(S)\}.$$

(The condition that μ is in $\Delta(L)$ means that, if $V \neq L$ then μ is a distribution in which no voters are choosing ballots outside of L. In the distribution μ, $\Sigma_{S \supseteq \{i\}} \mu(S)$ is the fraction of the voters who are including party i among the set of parties to which they give approval votes.) Then, in general, we may say that an electoral system (V,F) is an approval voting (AV) system iff,

(3) $$\Sigma_{i \in M(\mu)} F_i(\mu) = 1, \quad \forall \mu \in \Delta(L).$$

There are several ways to define approval-voting outcomes in the case of ties and complete the definition of the outcome function F. The simplest way is to let $V = L$ and specify that seats would be divided equally among the winning parties, and so

(4) $$F_i(\mu) = 1/|M(\mu)| \quad \text{if} \quad i \in M(\mu),$$
$$= 0 \quad \text{if} \quad i \notin M(\mu).$$

3. Coalitional straightforwardness.

In Section 1, we argued that barriers to entry can be created by voters' need to find preelection leadership to coordinate their votes. Preelection leadership is, by definition, not chosen in the election itself, and unelected leadership makes political change more difficult. Thus, we should seek electoral systems that minimize the need for such unelected leadership. In this section, we consider one way to formalize this idea of "minimizing the need for coordinating leadership."

The theorem of Gibbard [1973] and Satterthwaite [1975] implies that, if voters can have arbitrary preferences over legislative seat allocations, it is essentially impossible to design an electoral system in which voters can always identify their optimal voting strategies without any information about each other. So no electoral system can eliminate the incentive for voters to share information and coordinate in all situations. Thus, we focus here on the incentive to coordinate in a block of voters who have a very simple class of dichotomous preferences. That is, we consider here a block of voters who like all the parties in some set S, and who dislike all the other parties, and who simply want to maximize the fraction of the legislature that is allocated to the parties that they like.

To understand why we should focus on dichotomous preferences, consider first the situation in which a voter has completely identified his interests with some party j, and so he simply wants to maximize the number of seats that party j gets. If any voters do not need preelection leadership, then surely such a voter should not, because the leaders of party j are his leaders. Just knowing that he is for party j should be all that he needs to know on election day. There should be no need for a higher-level leader to tell such supporters of party j how to vote.

We are seeking electoral systems that minimize the deterrents against new parties that would adopt the same position as an existing party, because such deterrents would constitute barriers to entry that protect the politicians in the existing party. So suppose now that some other parties do adopt exactly the same position as party j, and let S denote the set of parties that adopt this position. That is, S is a set of parties that includes j, and a representative from any party in S would behave identically in the legislature. Then the voters who formerly identified with party j should now have the objective of maximizing the total number of legislative seats that are allocated to the parties in S. To minimize barriers to entry, there should be no need for these voters to follow a leader who transends the party structure and tells them whether to continue supporting party j or transfer support to other parties. Simply knowing that they are for the parties in S should be all that these voters need to know to identify their optimal ballots.

So consider a block (or set) of voters who constitute some fraction p of the total electorate, where $0 < p < 1$. Let μ denote the distribution of ballots submitted by the voters who are not in this block. If the distribution of ballots from the voters within the block is λ, then the distribution of ballots in the whole district will be $p\mu + (1 - p)\lambda$, and so the fraction of the legislative seats that are won by the parties in S will be

$$\Sigma_{i \in S} F_i((1 - p)\mu + p\lambda).$$

We may say that a ballot w is a <u>dominant vote for S</u> iff setting λ equal to $[w]$ maximizes this sum independently of μ and p. That is, w is a dominant vote for S iff

$$\Sigma_{i \in S} F_i((1 - p) + p[w]) \geq \Sigma_{i \in S} F_i((1 - p)\mu + p\lambda),$$
$$\forall p \in [0,1], \quad \forall \mu \in \Delta(V), \quad \forall \lambda \in \Delta(V).$$

(Recall that $[w]$ is the ballot distribution in which everyone submits w, so $(1 - p)\mu + p[w]$ is the ballot distribution that results if everyone in the block of size p submits the ballot w while μ is the distribution of ballots from the the other voters.)

Thus, our search for electoral systems that minimize the need for coordinating leadership brings us to look for electoral systems in which dominant votes exist. Relabelling the ballot set if necessary, we can assume without loss of generality that, if any ballot in V is a dominant vote for S then the ballot "S" itself is a dominant vote for S. That is, if there is a voting strategy that is always optimal for supporters of the parties in S, then one such voting strategy should be to simply write on the ballot the names of the parties in S.

So let us say that a voting system (V,F) is <u>coalitionally straightforward</u> iff $V \supseteq L$ and, for every S in L,

$$\Sigma_{i \in S} F_i((1 - p)\mu + p[S]) \geq \Sigma_{i \in S} F_i((1 - p)\mu + p\lambda),$$
$$\forall p \in [0,1], \quad \forall \mu \in \Delta(V), \quad \forall \lambda \in \Delta(V).$$

That is, an electoral system is coalitionally straightforward iff, for any coalition S such that $S \subseteq K$ and $\emptyset \neq S \neq K$, the ballot "S" is a dominant vote for S. (There is no need to worry about dominant votes for K, the set of all parties, because $\Sigma_{i \in K} F_i(\mu)$ always equals 1.)

To reinterpret coalitional straightforwardness, suppose that there is a block of voters who care only about some simple yes-or-no question of government policy (e.g. should our nation go to war?), and who simply want to maximize the number of seats occupied by legislators who will vote Yes on this question. If the electoral system is coalitionally straightforward then, no matter how many independent parties may advocate the Yes position and seek to represent this block of voters, the voters in this block will not need any preelection leadership to help them coordinate their votes with each other or the rest of the electorate. All that they need to know is the names of the Yes parties, because it is always optimal for them to simply list all these parties on their ballots. Thus, when several groups of political leaders are competing to represent a block of voters, this competition would not create a need for a higher level of leadership to tell the block how to vote, at least in this case where the voters in the block have the simplest possible preferences about government policy.

4. Nondiviseness.

A basic goal of the political process is to form a unifying government for a diverse electorate. This integrative function would be badly served if the electoral system positively encouraged schisms and discouraged consolidation of political organizations. Thus, to characterize good electoral systems, we also need an axiom of nondivisiveness. That is, in addition to asking that an electoral system should minimize barriers to entry against new political parties, we also want that an electoral system should minimize barriers to consolidation of existing political parties.

Given an electoral system (V,F), for any ballot distribution μ in $\Delta(V)$ and any number p between 0 and 1, let $C(\mu,p)$ denote the convex hull of the legislative seat allocations that a block could implement by putting all its support behind one party, when p is the fraction of the electorate in the block, and μ is the ballot distribution submitted by the voters who are not in the block. That is, let

$$C(\mu,p) = \{\Sigma_{i\in K} r_i F((1-p)\mu + p[\{i\}]) \mid r \in \Delta(K)\}.$$

Here $F((1-p)\mu + p[\{i\}])$ denotes the distribution of seats that would occur when the voters in the block of size p give their votes only to party i (that is, they all vote for the coalition {i}) and μ is the ballot distribution from the other voters. So when $r = (r_i)_{i\in K}$ is in $\Delta(K)$,

$$\Sigma_{i\in K} r_i F((1-p)\mu + p[\{i\}])$$

is a weighted average of seat distributions that the block could achieve by allocating their support to a single party.

We say that an electoral system is <u>nondivisive</u> iff the seat distribution that a block of voters could achieve by voting for any coalition of parties is always a weighted average of the seat distributions that a slightly larger block could achieve by consolidating its support behind a single party, when the ballot distribution from the voters outside the block remains the constant. That is, (V,F) is nondivisive iff, for every block size p and every number ε such that $0 < p < p + \varepsilon < 1$,

$$F((1-p)\mu + p[S]) \in C(\mu,p + \varepsilon), \quad \forall\mu \in \Delta(V), \quad \forall S \in L.$$

According to this definition, if the electoral system is nondivisive then, in any direction that one might want to move the distribution of seats, a block of voters that spreads its support among many parties should not be able to move the seat distribution any farther than a slightly larger block that gives all its support to one well-chosen party. So nondivisiveness asserts that the only essential reasons for a block to vote for a coalition of several parties must be uncertainty (about the distribution of votes outside the block, or about the size of the block itself) or nonlinearity of preferences on $\Delta(K)$

(e.g., a desire to not let any one party get too big).

Proportional representation satisfies both the coalitional straightforwardness and the nondivisiveness properties as formulated here, whereas plurality voting satisfies nondivisiveness but violates coalitional straightforwardness. Thus, we may infer that proportional representation has both low barriers to entry and low barriers to consolidation, but plurality voting has high barriers to entry with low barriers to consolidation. We may expect that an electoral system with low barriers to both entry and consolidation would be essentially neutral with respect to the number of political parties that form; and so the number of political parties could be large or small, in response to the social and cultural needs of the society. In contrast, an electoral system with high barriers to entry but low barriers to consolidation would tend to generate small numbers of political parties. Analysis of empirical data (see Rae [1971] and Lijphardt [1990]) tends to confirm these predictions for proportional representation and plurality voting. Countries with proportional representation have widely varying numbers of political parties, but countries with plurality voting rarely have more than three parties. The observed fact that plurality voting (in single-member districts) generally leads to the formation of just two political parties is known as Duverger's Law. (See Riker [1982b].)

Approval voting also satisfies both coalitional straightforwardness and nondivisivenss conditions. However, approval voting would violate nondivisiveness if we altered the definition of nondivisiveness by eliminating the ε parameter or by allowing it to equal zero. To see why, consider a three-party example with $K = \{1,2,3\}$, and consider the simple symmetric tie-breaking rule discussed at the end of Section 2. (Actually, any version of approval voting that satisfies the axioms of this paper will do.) Suppose that 1/3 of the electorate votes for the set {1,2}, 1/3 of the electorate votes for {1,3}, and 1/3 of the electorate votes for {2,3}. Then each party is getting approval votes from 2/3 of the voters, and so the three parties divide the seats equally in the seat distribution (1/3, 1/3, 1/3). But if the block of voters who are voting {1,2} changed their votes to support only party 1, then the seat distribution would change to (1/2, 0, 1/2), because only parties 1 and 3 would be tied for most approval votes. Similarly, if the block of voters who are voting {1,2} changed their votes to support only party 2 then the seat distribution would change to (0, 1/2, 1/2); and if this block changed their votes to support only party 3 then the seat distribution would change to (0, 0, 1). Thus, when this block consolidates their support behind just one party, they cannot reduce party 3's seat allocation below 1/2, but they can reduce party 3's seat allocation to 1/3 by supporting both parties 1 and 2. Thus, this block may seem to derive some advantage from having two different parties to support.

However, this advantage is not robust to even tiny changes in the size of the block. If the block were enlarged by even one voter, keeping the number and ballots of the other voters fixed, then the members of this block could generate the allocation (1, 0, 0) by voting for party 1, and they could generate the allocation (0, 1, 0) by voting for party 2. Thus, the advantage that this block derives from having two parties to support is not more than the advantage that the block could derive from recruiting just one more voter to join them.

Although proportional representation in its ideal form satisfies both coalitional straightforwardness and nondivisiveness, these properties are not satisfied by the finite approximations to proportional representation that are used in districts with a finite number of indivisible seats. For example, consider a district with 2 indivisible legislative seats, and suppose that, if each voter endorsed a single party, then these two seats would be allocated according to the greatest-divisor method of Jefferson and d'Hondt. (See Balinski and Young [1982].) Consider a situation in which 30% of the electorate is expected to vote for party 1, 25% is expected to vote for party 2, 25% is expected to vote for party 3, and the remaining 20% of the electorate is a block of voters who want to maximize the total number of seats for parties 2 and 3. If this 20% block concentrated all their support on any single party, then party 1 would get at least one of the two seats. For example, if everyone in the block voted for party 2, then party 2 would earn a seat with 45% of the votes, party 1 would earn a seat with 30% of the votes, and party 3 with 25% would get no seats. (Notice also that increasing the block slightly would not change this result.) However, if the block divided its support equally between parties 2 and 3, then parties 2 and 3 would each earn a seat with 35% of the votes, while party 1 with 30% would not get a seat.

This example shows that our nondivisiveness property is stronger than the property of "encouraging coalitions" that has been discussed by Balinski and Young [1982] as a motivation for the greatest-divisor method. For an electoral system to be nondivisive in our sense, it is not enough that a set of parties would not lose seats if the leaders of these parties decided to merge them. We require here that any small block of voters who favor these parties should be able to initiate the consolidation process, by concentrating their support behind one party, without reducing the total number of seats that these parties get.

5. Neutrality, responsiveness, and homogeneity.

In free democracies, the rules of the electoral system should not intrinsically favor any one party. So the outcome of the election should not depend on the party labels. A relabelling of the parties is any function $\pi:K \to K$ that is one-to-one and onto. (That is, $\pi(i) \neq \pi(j)$ if $i \neq j$.) For any relabelling π and any set of parties S, let

$$\pi(S) = \{\pi(i) | i \in S\}.$$

Then we may say that an electoral system (V,F) is neutral iff, for each relabelling $\pi:K \to K$, for each distribution of coalitional ballots $r = (r_S)_{S\in L}$ in $\Delta(L)$, and for each party i in K,

$$F_i(\Sigma_{S\in L}\ r_S[S]) = F_{\pi(i)}(\Sigma_{S\in L}\ r_S[\pi(S)]).$$

That is, an electoral system is neutral if the outcome would not be changed by permuting the names of the parties, at least in the case when voters all cast coalitional ballots.

None of our assumptions has as yet ruled out the possibility that the electoral system simply divides the seats equally among the parties no matter how people vote. To rule out this system, we need some assumption that says that control of the seat distribution really does belong to the voters. One natural axiom of this form is to assert that, if all voters unanimously vote for one unique party, then that party should get all the seats. In our notation, $\{i\}$ denotes the simple vote for party i, and $[\{i\}]$ denotes the ballot distribution in which everyone submits a simple vote for party i. So we may say that an electoral system (V,F) is responsive iff

$$F_i([\{i\}]) = 1, \quad \forall i \in K.$$

Approval voting and proportional representation satisfy all of the conditions that we have developed so far: coalitional straightforwardness, nondivisiveness, neutrality, and responsiveness. Other electoral systems that satisfy these four conditions can be constructed by dividing the seats into two groups, of which one group is allocated by approval voting and the other group by proportional representation. At the time of this writing, I do not know of any electoral systems that satisfy these four conditions other than mixtures of PR and AV, but I also cannot disprove the existence of other electoral systems that satisfy these conditions. The characterization that I do have requires one more condition that is motivated by a political interpretation which may be less compelling than neutrality and responsiveness.

Suppose, for example, that the parties in S are leftist parties and the other parties in K\S are rightist parties. Adding a block of leftist voters who vote for S obviously might decrease the share of seats that are allocated to the rightist parties. But we may want to stipulate that, when voters in some block choose to give maximal support to the leftist parties, they should not have any say about how the rightist parties proportionally divide among themselves the seats (if any) that they win in this district. That is, to provide some autonomy for the rightist faction, we may ask that the relative importance of the various rightist parties should be determined only by the rightist voters themselves, not by the leftist voters.

Adding a block of votes for S would change a ballot distribution μ to a distribution of the form $(1 - p)\mu + p[S]$, where p is between 0 and 1. To assert that the parties outside of S would not change their relative strengths when such a block of votes for S is added, we may want to write

$$F_i((1-p)\mu + p[S])/F_j((1-p)\mu + p[S]) = F_i(\mu)/F_j(\mu), \quad \forall i \in K\backslash S, \quad \forall j \in K\backslash S.$$

Unfortunately, this homogeneity equation is not well-defined if the denominators are zero. However, we can equivalently (but less intuitively) rewrite this condition by multiplying both sides by the product of the denominators. Thus, we may say that an electoral system (V,F) is <u>homogeneous</u> (or <u>coalitionally autonomous</u>) iff,

$$F_i((1-p)\mu + p[S])\ F_j(\mu) = F_j((1-p)\mu + p[S])\ F_i(\mu),$$
$$\forall S \in L, \quad \forall p \in [0,1], \quad \forall \mu \in \Delta(V), \quad \forall i \in K\backslash S, \quad \forall j \in K\backslash S.$$

(Equivalently, we may say that (V,F) is homogeneous iff, for any S in L, for any p between 0 and 1, and for any μ in $\Delta(V)$, there exists some number q such that, for every party i that is not in S,

$$F_i((1-p)\mu + p[S]) = qF_i(\mu).$$

Here q may depend on μ, p, and S, but q does not depend on $i \in K\backslash S$.)

6. Proportional representation and approval voting.

It is not difficult to verify that simple examples of approval voting and proportional representation that we formulated above, in equations (2) and (4) in Section 2 (with V = L), satisfy all of the five conditions that we have developed in Sections 3 through 5. In fact, all other electoral systems that satisfy these five conditions are just generalizations of these two examples. This result is stated in the following theorem, which is proven in Section 9.

<u>Theorem</u>. Suppose that (V,F) is coalitionally straightforward, nondivisive, neutral, responsive, and homogeneous. Then (V,F) is either an approval voting system or a proportional representation system (as defined by equations (1) and (3) in Section 2).

To appreciate power of this theorem, it may be helpful to consider some other electoral systems that do not satisfy all five of the conditions. Plurality voting violates only coalitional straightforwardness, among these five conditions. Mixed systems in which some fraction of the seats are allocated by AV and some by PR, violate only the homogeneity condition.

The term _approval voting_ is used here to mean "multiple approval votes, winner take all," whereas the term _proportional representation_ is used here to mean "single divisible vote, proportional seat allocation." In contrast, we might consider a "multiple approval votes, proportional seat allocation" electoral system, in which each voter can give approval votes to as many parties as he wants, and legislative seats are allocated in proportion to the number of approval votes that each party gets. To formalize such an electoral system, let $V = L$ and

$$F_i(\mu) = (\Sigma_{S\supseteq\{i\}} \mu(S))/(\Sigma_{j\in K} \Sigma_{T\supseteq\{j\}} \mu(T)), \quad \forall\mu \in \Delta(L), \quad \forall i \in K.$$

This system satisfies four of our five conditions, but it violates nondivisiveness. This violation of nondivisiveness is indeed very problematic, and makes this system a bad way to elect a legislature. For example, suppose that 60% of the electorate are leftists and 40% are rightists, but there are two rightist parties competing with only one leftist party. Then the single leftist party would get 6/14 of the seats whereas each rightist party would get 4/14 of the seats, giving the rightists an overall 8/14 majority.

The conclusion of the above theorem is weakened by the fact that our definitions of AV systems and PR systems (in Section 2) say nothing about the outcomes that may occur when some voters submit ballots that do not simply list a coalition of parties (as may occur when V is strictly larger than L). However, the coalitional straightforwardness axiom imposes significant restrictions on the effect of noncoalitional votes, because noncoalitional votes cannot be better than coalitional votes for voters with dichotomous preferences, and these restrictions have not been formulated in this theorem,. So there seems to be a need for future research to find a stronger conclusion in this theorem.

Of course, other ways to formulate the ideas of "low barriers to entry" and "low barriers to consolidation" (different from our coalitional straightforwardness and nondivisiveness conditions) should be considered in future research. It seems reasonable to expect that other axiomatic formulations of these ideas may be found that lead similarly to approval voting and proportional representation systems. One might, for example, consider axioms that pertain to changes in equilibrium outcomes when parties merge or split. Such axioms have been avoided in this paper because they require an agreed-upon equilibrium concept for elections, and because changing the set of parties K means changing the range of the outcome function $F:\Delta(V) \to \Delta(K)$ that is used here as the basic model of an electoral system. Thus, axioms about the effects of splitting and merging parties would require a more complicated model of an electoral system than we have used here. (See also the work of Gibbard [1978] on random social choice mechanisms, which can be reinterpreted as an axiomatization of proportional representation. The main theorem of Gibbard [1978] implies that, if an electoral system is responsive and is straightforward for every voter whose objective is to maximize a linear function of the seat distribution, then this electoral system must be a proportional representation system.)

7. Other factors affecting barriers to entry.

The main focus in this paper is on the barriers to entry that may be created by the rules of the electoral system. We must note, however, that barriers to entry in political systems depend on much more than just the rules of the electoral system itself.

Freedom of speech is clearly the first essential prerequisite for lowering barriers to entry in politics. Conversely, economies of scale in the mass media, and voters' limited ability to assimilate political information, can raise barriers to entry and put an upper bound on the number of parties that can effectively compete for voters' support, regardless of the electoral system.

Regional autonomy is another general factor that can help reduce barriers to entry. Locally and regionally elected offices create opportunities for politicians to develop independent local power bases, which can allow them to challenge established national leaders. Local and regional administrative offices (if they are not subservient to centrally appointed prefects) create opportunities for local politicians to prove their ability to govern, and to develop the credentials that they need to offer themselves as alternatives for national leadership.

The concept of barriers to entry into national leadership can also help us to understand why regional districts may be important even in PR elections of a national legislature. If the entire country is one PR district, then each party must nominate a single national list, which must be negotiated by the party's central committee. So a legislator who wants to be renominated cannot seriously challenge the leadership of his party unless he is prepared to leave it and create a party of his own. On the other hand, when the PR election is broken up into regional districts, then regional party organizations can take some control over the nomination process. With regionally decentralized nomination, legislators who have strong support in their regions gain some independence to challenge the national party leadership.

8. Constitutional implications.

The main result of this paper is that proportional representation and approval voting are uniquely compatible, in terms of several important properties that they share. In particular, PR and AV both reduce barriers to entry for new parties without being divisive, and thus they both allow flexible party configurations that respond to social and political needs. This result suggests that, in the broader questions of constitutional design, we should consider structures that mix these two systems.

With respect to incentives for party positioning, however, proportional representation and approval voting have very different and complementary properties. Theoretical analysis suggests that parties that advocate centrist positions would be rewarded under approval voting, perhaps even more than under plurality voting. (For example, see Brams and Fishburn [1983], Cox [1985], and Myerson and Weber [1988].) In contrast, proportional representation encourages parties to adopt a broad range of positions. (See Austin-Smith and Banks [1988], for example).

The relative desirability of centrism and breadth in a legislature should be discussed with reference to two rival theories of where government policy should be determined: in the general election, or in legislative bargaining.

Of course, details of policy determination must be left to bargaining among elected representatives, because general public debate on every detail is impossible. However, the extreme multiplicity of equilibria in many natural bargaining models (see Myerson [1991], chapter 8) suggests that trying to resolve everything by bargaining among representatives who perfectly reflect the whole spectrum of public opinion may lead to a legislature that is unpredictable, arbitrary, or deadlocked. The problem is that, because legislators do not act anonymously, their actions may be distorted by a desire to cultivate (or avoid) a reputation that would be helpful (or harmful) in future bargaining. For example, a legislative leader may fear that, if he offers concessions beyond some arbitrary point and agrees to support a bill that seems less than ideal to him, then he may be expected to offer many more costly concessions in future legislative bargaining. On the other hand, voters in a general election submit their ballots anonymously and so have fewer such reputational incentives. This distinction may be an important reason to encourage electoral policy determination before legislative policy determination, whenever possible. (For the similar reasons, economists generally recommend that prices should be determined by anonymous market forces, rather than by bargaining between representatives of a buyers' organization and a sellers' cartel.)

In comparisons between political systems, we may say that electoral policy determination increases as the variance of policy positions that are advocated by the elected representatives decreases, when the distribution of voters' policy preferences is held fixed. Thus, the goal of maximizing electoral policy determination implies that an electoral system should be <u>centrist</u>, in the sense that it tends to give the most legislative seats to the parties that advocate compromise positions which can get wide support among voters.

On the other hand, one might also ask that a legislature should be broad, in the sense that it should include representatives of all major ideological and geographical factions. If a legislature is not broad, then significant segments of public opinion may feel left out of policy debates. Conversely, when a national consensus is needed to cope with a major crisis, this consensus can be expressed in the legislature only if it is sufficiently broad.

Breadth and centrism are not necessarily incompatible, because inclusion of all major factions does not necessarily imply that these factions all get seats in proportion to their support in the electorate. One way to create a legislature that is both broad and centrist is by combining PR and AV in a bicameral legislature, thus making use of both their compatibility and their differences. Thus, let us consider a legislature in which the lower house is elected by proportion representation, and the upper house consists of representatives from single-member districts who are elected by approval voting. In such a legislature, we may expect to find breadth in the PR house, and strong centrism in the AV house.

Having two such different houses could make it hard to pass legislation, especially because the formation of a coherent majority is often difficult in a broadly representative PR house, which may include a wide range of idiosyncratic parties. One way to avoid such indecisiveness is to lower the quota for passing bills through the PR house. That is, the constitution could specify that, to become law, a bill needs to be approved by more than half of the upper AV house and by more than one-third of the lower PR house. A one-third quota is the lowest quota such that at most two disjoint coalitions can pass bills out of the lower house. Notice that, to prevent the legislature from passing laws that contradict each other without any legislator changing his or her mind, a majority quota is really only needed in one house. With its lower quota, the broad lower house could take a leading role in formulating bills, while the centrist upper house with its higher quota would get a more important role in ratifying bills. Because two disjoint coalitions could pass bills out of the lower house, these legislative rules would allow an opposition coalition to create a tangible record of their versions of all major legislation.

(People have relatively little experience with procedural rules for a less-than-majority quota, so some possible procedural rules for such a legislature might be worth discussing here. Suppose for now that bills ordinarily start in the lower house. When a bill is proposed before the lower house, alternative versions that are sponsored by two or more representatives can all be brought up simultaneously for a primary lower-house vote, by approval voting. If two or less versions get the 1/3+ quota of the lower house, then these versions are sent to the upper house. If more than two versions get the 1/3+ quota in the primary vote, then these versions of the bill should be reconsidered in a second lower-house vote in which each representative can only vote for one version. Any versions that get the 1/3+ quota on this second vote are sent to the upper house. Notice that at most two such versions will be sent. If two versions are sent to the upper house, then there is a primary upper-house vote where a majority of those voting determines which version goes to the final upper-house vote. In the final vote, a majority of the upper house is needed to pass the bill, in its one remaining version, into law.)

Other ways to combine approval voting and proportional representation in constitutional design may be worth considering. For example, constitutional debates in some countries have recently focused on proposals to combine a PR legislature with a directly elected president or chief executive. (See Susser [1989].) Our results suggest that, in such a constitution, approval voting could be recommended as a procedure for electing the president, to make the presidential elections compatible with the party structure that is induced by PR in the legislative elections.

Of course, the formal properties that we have studied here are only imperfect and incomplete formulations of the criteria by which a political scientist might try to identify better constitutional structures. We may anticipate that further axiomatic analysis of electoral systems and other constitutional provisions may be a rich source of new and practical ideas about how to design political institutions that can improve the chances for a successful functioning democracy.

9. Proof of the Theorem.

Throughout this proof, we assume that $F:\Delta(V) \to \Delta(K)$ is coalitionally straightforward, nondivisive, neutral, responsive, and homogeneous. Recall also that we assume $|K| \geq 3$ and $V \supseteq L$, where L is the set of all proper subsets of K.

Homogeneity and neutrality imply that, if each voter names a single party on his ballot, then the ratio of seats that two parties get should depend only on the ratio of the number of votes that these parties get. That is, there must exist some function $\phi:[0,1] \to [0,1]$ such that, for any r in $\Delta(K)$, for any two parties h and j, if $0 \leq r_h \leq r_j$ and $0 < r_j$ then

$$F_h(\Sigma_{i\in K}\ r_i[\{i\}]) = \phi(r_h/r_j)\ F_j(\Sigma_{i\in K}\ r_i[\{i\}]).$$

(Here, $[0,1]$ is the set of all numbers between 0 and 1.) To prove this fact, given any two different parties h and j, let

$$\phi(q) = \frac{F_h((q/(1+q))[\{h\}] + (1/(1+q))[\{j\}])}{F_j((q/(1+q))[\{h\}] + (1/(1+q))[\{j\}])}.$$

By neutrality, this function is independent of which two parties h and j are considered. Homogeneity implies that mixing in any distribution of votes for parties other than h and j cannot affect the ratio of the numbers of seats that go to parties h and j.

Neutrality and responsiveness also imply that $\phi(1) = 1$ and $\phi(0) = 0$. By coalitional straightforwardness, the function ϕ must be nondecreasing. The assumption that K includes at least three parties implies that ϕ must also satisfy the equation $\phi(pq) = \phi(p)\phi(q)$, for all numbers p and q in $[0,1]$. A nondecreasing function ϕ can satisfy these conditions only if there exists some number α, which we may call the <u>scoring exponent</u>, such that

$$0 \le \alpha \le +\infty \quad \text{and} \quad \phi(p) = p^{\alpha} \quad \forall p \in [0,1].$$

(Here we let $p^{+\infty} = 0$ if $0 \le p < 1$, $1^{+\infty} = 1$, and $0^0 = 0$.) To prove this claim, let α be such that $\phi(.5) = .5^{\alpha}$. Let n and m be any integers. Because $\phi(.5^{n/m})^m = \phi(.5^n) = \phi(.5)^n = .5^{\alpha n}$, we must have $\phi(.5^{n/m}) = (.5^{n/m})^{\alpha}$. That is, $\phi(p) = p^{\alpha}$ for any number p that is a nonnegative rational power of 0.5. Then the formula $\phi(p) = p^{\alpha}$ can be extended to all p in $[0,1]$ because ϕ is nondecreasing.

Thus, when each voter supports one party, the seats must be allocated in proportion to the α-power of the number of supporters. That is, if $\alpha < +\infty$ then, for any r in $\Delta(K)$,

$$F_j(\Sigma_{i\in K}\ r_i[\{i\}]) = (r_j)^{\alpha}/(\Sigma_{i\in K}\ (r_i)^{\alpha}).$$

If $\alpha = +\infty$ (which occurs when $\phi(.5) = 0$) then, for any r in $\Delta(K)$,

$$\{j \mid F_j(\Sigma_{i\in K} r_i[\{i\}]) > 0\} \subseteq \text{argmax}_{j\in K} r_j.$$

That is, when α is infinite, all seats go to the parties with most support.

If α were less than 1, then support for parties would give decreasing returns, which implies a violation of nondivisiveness. For example, let $K = \{1,2,3\}$. Then $\alpha < 1$ would imply that

$$\Sigma_{j\in\{1,2\}} F_j(.25[\{1\}]+.25[\{2\}]+.5[\{3\}])$$

would be strictly greater than 1/2 and strictly greater than $\Sigma_{j\in\{1,2\}} F_j([\{3\}])$ and $\Sigma_{j\in\{1,2\}} F_j((.5+\varepsilon)[\{i\}]+(.5-\varepsilon)[\{3\}])$, for all i in $\{1,2\}$ and all sufficiently small positive ε. But $F(.5[\{1,2\}] + .5[\{3\}])$ is in the convex hull of

$$\{F([\{3\}]),\ F((.5+\varepsilon)[\{1\}]+(.5-\varepsilon)[\{3\}]),\ F((.5+\varepsilon)[\{2\}]+(.5-\varepsilon)[\{3\}])\}$$

by nondivisiveness, and

$$\Sigma_{j\in\{1,2\}} F_j(.5[\{1,2\}] + .5[\{3\}])$$
$$\geq \Sigma_{j\in\{1,2\}} F_j(.25[\{1\}]+.25[\{2\}]+.5[\{3\}]),$$

by coalitional straightforwardness. Thus, the scoring exponent α must satisfy $1 \leq \alpha \leq +\infty$.

Lemma. Given any μ in $\Delta(V)$, any S in L, and any p in [0,1], suppose that there exists some h in K such that

$$\Sigma_{i\in S} F_i((1-p)\mu + p[\{h\}])$$
$$> \Sigma_{i\in S} F_i((1-p)\mu + p[\{j\}]), \quad \forall j \in K\backslash\{h\},$$

and, for every j in K, $F((1-p-\varepsilon)\mu + (p+\varepsilon)[\{j\}])$ is a continuous function of ε in some interval $[0,e]$, where $0 < e < 1$. Then

$$F((1-p)\mu + p[S]) = F((1-p)\mu + p[\{h\}]).$$

The proof of this lemma follows immediately from nondivisiveness and coalitional straightforwardness. Nondivisiveness implies that $F((1-p)\mu + p[S])$ is in $C(\mu,p+\varepsilon)$ for every positive ε. The continuity condition in the theorem implies that any sequence of points in $C(\mu,p+\varepsilon)$ must approach $C(\mu,p)$ as $\varepsilon \to 0$. So $F((1-p)\mu + p[S]) \in C(\mu,p)$. Coalitional straightforwardness implies that

$$\Sigma_{i\in S} F_i((1-p)\mu + p[S]) \geq \Sigma_{i\in S} F_i((1-p)\mu + p[\{h\}]).$$

But the assumptions in the lemma specify that $F((1-p)\mu + p[\{h\}])$ uniquely maximizes the total allocation of seats to parties in S, among all allocations in $C(\mu,p)$. So the conclusion of the lemma follows.

By analyzing a specific example, we now show that α must be either 1 or $+\infty$. So suppose now (contrary to this claim) that $1 < \alpha < +\infty$. Using neutrality and the existence of at least three parties, we may without loss of generality assume that $K \supseteq \{1,2,3\}$. When α is finite,

$$F(r_1[\{1\}] + r_2[\{2\}] + r_3[\{3\}])$$

is a continuous function of (r_1,r_2,r_3) in $\Delta(\{1,2,3\})$. When $\alpha > 1$,

$$\begin{aligned}&\Sigma_{i\in\{1,2\}} F_i(.52[\{1\}] + .48[\{3\}])\\ &\quad > \Sigma_{i\in\{1,2\}} F_i(.50[\{1\}] + .02[\{2\}] + .48[\{3\}])\\ &\quad > \Sigma_{i\in\{1,2\}} F_i(.50[\{1\}] + .5[\{3\}]).\end{aligned}$$

That is, $\alpha > 1$ implies that a block of size .02 that wants to maximize the total number of seats allocated to parties 1 and 2 can maximize its impact by voting for party 1 if, among the other voters, there are more votes for party 1 than for party 2. In effect, $\alpha > 1$ implies that there are increasing returns to joining other voters in support of a party. So, by the lemma,

$$F(.50[\{1\}] + .02[\{1,2\}] + .48[\{3\}]) = F(.52[\{1\}] + .48[\{3\}]).$$

Similarly,

$$\Sigma_{i\in\{1,2\}} F_i(.52[\{1\}] + .47[\{2\}] + .01[\{3\}])$$
$$> \Sigma_{i\in\{1,2\}} F_i(.50[\{1\}] + .49[\{2\}] + .01[\{3\}])$$
$$> \Sigma_{i\in\{1,2\}} F_i(.50[\{1\}] + .47[\{2\}] + .03[\{3\}])$$

and so

$$F(.50[\{1\}] + .02[\{1,2\}] + .47[\{2\}] + .01[\{3\}])$$
$$= F(.52[\{1\}] + .47[\{2\}] + .01[\{3\}]).$$

Furthermore, when $\alpha > 1$,

$$\Sigma_{j\in\{2,3\}} F_j(.52[\{1\}] + .48[\{3\}])$$
$$> \Sigma_{j\in\{2,3\}} F_j(.52[\{1\}] + .47[\{2\}] + .01[\{3\}])$$
$$> \Sigma_{j\in\{2,3\}} F_j(.99[\{1\}] + .01[\{3\}])$$
$$= \Sigma_{j\in\{2,3\}} F_j(.97[\{1\}] + .02[\{1,2\}] + .01[\{3\}]).$$

These results imply that

$$F(.50[\{1\}] + .02[\{1,2\}] + .47[\{2,3\}] + .01[\{3\}])$$
$$= F(.52[\{1\}] + .48[\{3\}]).$$

On the other hand,

$$\Sigma_{j\in\{2,3\}} F_j(.50[\{1\}] + .49[\{2\}] + .01[\{3\}])$$
$$> \Sigma_{j\in\{2,3\}} F_j(.50[\{1\}] + .02[\{2\}] + .48[\{3\}])$$
$$> \Sigma_{j\in\{2,3\}} F_j(.97[\{1\}] + .02[\{2\}] + .01[\{3\}])$$

and so

$$F(.50[\{1\}] + .02[\{2\}] + .47[\{2,3\}] + .01[\{3\}])$$
$$= F(.50[\{1\}] + .49[\{2\}] + .01[\{3\}]).$$

Then using coalitional straightforwardness, we conclude that

$$\Sigma_{i\in\{1,2\}} F_i(.52[\{1\}] + .48[\{3\}]).$$
$$= \Sigma_{i\in\{1,2\}} F_i(.50[\{1\}] + .02[\{1,2\}] + .47[\{2,3\}] + .01[\{3\}])$$
$$\geq \Sigma_{i\in\{1,2\}} F_i(.50[\{1\}] + .02[\{2\}] + .47[\{2,3\}] + .01[\{3\}])$$
$$= \Sigma_{i\in\{1,2\}} F_i(.50[\{1\}] + .49[\{2\}] + .01[\{3\}]).$$

That is, we must have

$$.52^{\alpha}/(.52^{\alpha} + .48^{\alpha}) \geq (.50^{\alpha} + .49^{\alpha})/(.50^{\alpha} + .49^{\alpha} + .01^{\alpha}).$$

But this inequality does not hold for any α such that $1 < \alpha < +\infty$. (It would hold only if $(52/48)^{\alpha} \geq 50^{\alpha} + 49^{\alpha}$, which is certainly not true for any finite positive α.)

Thus, the scoring exponent α cannot be strictly between 1 and $+\infty$. That is, we must have either $\alpha = 1$ or $\alpha = +\infty$. When $\alpha = 1$, we have a proportional representation system, as defined in Section 2. So let us henceforth consider the case of $\alpha = +\infty$.

Given any μ in $\Delta(L)$, let $M(\mu) = \text{argmax}_{i \in K} \Sigma_{T \supseteq \{i\}} \mu(T)$. By induction in m, we can show that, for any vote distribution μ in $\Delta(L)$ such that m is the number of coalitions containing two or more parties that are endorsed by a positive fraction of the voters, the only parties that get seats are those in the set $M(\mu)$. For $m = 0$, the claim holds because $\alpha = +\infty$. Now suppose inductively that $m \geq 1$ and the claim holds for m-1.

Let μ be any vote distribution such that exactly m coalitions containing two or more parties are endorsed by a positive fraction of the voters. Let S be any coalition in L such that $\mu(S) > 0$ and S contains two or more parties. If $\mu(S) = 1$ then responsiveness and coalitional straightforwardness would together imply that $\Sigma_{i \in S} F_i(\mu) = 1$, which would prove the claim (because $S = M(\mu)$ when $\mu(S) = 1$). So we may suppose that $\mu(S) < 1$. Let η denote the distribution of votes among the voters who do not vote for S. That is,

$$\eta(S) = 0 \quad \text{and} \quad \mu = (1 - \mu(S))\eta + \mu(S)[S].$$

There would be only m-1 coalitions of two or more parties that are endorsed by a positive fraction of the voters if the S-supporters in μ switched to endorse a single party. Let

$$M^*(\mu,S) = \cup_{i\in K} M((1-\mu(S))\eta + \mu(S)[\{i\}])\}.$$

Notice that $i \in M^*(\mu,S)$ if and only if $i \in M((1-\mu(S))\eta + \mu(S)[\{i\}])$. A sufficiently small increase in $\mu(S)$ (with small proportional decrease in the other elements of μ) would not change $M^*(\mu,S)$, so nondivisiveness and the induction hypothesis imply that

$$\Sigma_{i\in M^*(\mu,S)} F_i(\mu) = 1.$$

To complete the proof, we consider two cases which together include all possibilities.

(Case 1.) First suppose that $M(\mu)\backslash S \neq \emptyset$. Then

$$M^*(\mu,S)\backslash M(\mu) \subseteq K\backslash S,$$

because the S-supporters cannot make a party in S get more approval votes than a party in $M(\mu)\backslash S$ by reassigning their votes from all of S to any single party. Furthermore, $M(\eta) = M(\mu)\backslash S$, so removing the S-supporters altogether would leave a vote distribution in which the parties in $M(\mu)\backslash S$ would get all the seats. That is, by the induction hypothesis for m-1,

$$\Sigma_{i\in M(\mu)\backslash S} F_i(\eta) = 1.$$

By homogeneity, going from η to μ cannot change the relative distribution of seats among the parties in $K\backslash S$, so

$$F_j(\mu) = 0, \quad \forall j \in (K\backslash S)\backslash M(\mu).$$

But $(K\backslash S)\backslash M(\mu) \supseteq M^*(\mu,S)\backslash M(\mu)$, so we can conclude that

$$\Sigma_{i\in M(\mu)} F_i(\mu) = 1.$$

(Case 2.) Now suppose instead that $M(\mu) \subseteq S$. By switching to support any one party in $M(\mu)$, the supporters of S could create a vote distribution in which, by the induction hypothesis, this party would get all the seats. Thus, by coalitional straightforwardness,

$$\Sigma_{i \in S} F_i(\mu) = 1.$$

If the theorem fails, then there must exist some party j such that $j \in S \backslash M(\mu)$ but $F_j(\mu) > 0$. Let h be any party that is not in S. (Here is the only place where we use the assumption that the set of all parties K is not in L.) Let

$$\mu' = (1-p)\mu + p[\{h\}],$$

where the number p is chosen so that

$$\max_{i \in M(\mu)} \Sigma_{T \supseteq \{i\}} \mu'(T) > \Sigma_{T \supseteq \{h\}} \mu'(T) > \Sigma_{T \supseteq \{j\}} \mu'(T).$$

Then $M(\mu') = M(\mu)$, and so the same argument implies that

$$\Sigma_{i \in S} F_i(\mu') = 1.$$

However, if we changed μ' by having the voters who support S in μ' switch to only support j or any other single party, we could not create a vote-distribution in which party j would have the most approval votes. That is, $j \notin M^*(\mu', S)$. This property would hold even if we perturbed μ' by a sufficiently small increase in the size of the block supporting S and correspondingly small proportional decreases in the sizes of all other voting blocks. Thus, by nondivisiveness and the induction hypothesis for m-1, $F_j(\mu') = 0$. By homogeneity, the change from μ to μ' cannot transfer seats from j to other parties in S, because $\mu' = (1-p)\mu + p[\{h\}]$ and $h \notin S$. So $F_j(\mu) = 0$. This contradiction of the way that we chose j implies that

$$\Sigma_{i \in M(\mu)} F_i(\mu) = 1.$$

Q.E.D.

REFERENCES

D. Austin-Smith and J. Banks [1988], "Elections, Coalitions, and Legislative Outcomes," American Political Science Review, 82, 405-422.

M. L. Balinski and H. P. Young [1982], Fair Representation, New Haven: Yale University Press.

W. J. Baumol, J. C. Panzar, and R. D. Willig, [1986], "On the Theory of Perfectly-Contestable Markets," in New Developments in the Analysis of Market Structure, edited by J. E. Stiglitz and G. F. Mathewson, Cambridge: M.I.T. Press.

S. J. Brams and P. C. Fishburn [1983], Approval Voting, Boston: Birkhauser.

S. J. Brams and P. C. Fishburn [1990], "Coalition Voting," New York University discussion paper.

G. W. Cox [1985], "Electoral Equilibrium under Approval Voting," American Journal of Political Science 29, 112-118.

A. Gibbard [1973], "Manipulation of Voting Schemes: A General Result," Econometrica 41, 587-601.

A. Gibbard [1978], "Straightforwardness of Game Forms with Lotteries as Outcomes," Econometrica 46, 595-614.

A. Lijphardt [1990], "The Political Consequences of Electoral Laws, 1945-85," American Political Science Review 84, 481-496.

R. B. Myerson [1991], Game Theory; Analysis of Conflict, Cambridge, MA: Harvard University Press (to appear).

R. B. Myerson and R. J. Weber [1988], "A Theory of Voting Equilibria," Northwestern University discussion paper.

D. W. Rae [1971], The Political Consequences of Electoral Laws, New Haven: Yale University Press.

W. H. Riker [1982a], Liberalism Against Populism, San Fransisco: Freeman.

W. H. Riker [1982b] "The Two-Party System and Duverger's Law," American Political Science Review 76, 753-766.

M. A. Satterthwaite [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.

H. A. Simon [1954], "Bandwagon and Underdog Effects and the Possibility of Equilibrium Predictions," Public Opinion Quarterly 18, 245-253.

J. Smith [1973], "Aggregation of Preferences with a Variable Electorate," Econometrica 41, 1027-1041.

B. Susser [1989], "'Parliadential' Politics: A Proposed Constitution for Israel," Parliamentary Affairs, 42, 112-122.

H. P. Young [1974], "Social Choice Scoring Functions," SIAM Journal of Applied Mathematics 28, 824-838.