

replaces California Institute of Technology  
Social Science Working Paper #688

JA  
74  
M63  
1989

# Models of Strategic Choice in Politics

HUN

Edited by Peter C. Ordeshook

*Ann Arbor*  
*The University of Michigan Press*

Copyright © by The University of Michigan 1989  
All rights reserved  
Published in the United States of America by  
The University of Michigan Press  
Manufactured in the United States of America

1992 1991 1990 1989 4 3 2 1

**Library of Congress Cataloging-in-Publication Data**

Models of strategic choice in politics / edited by Peter C. Ordeshook.

p. cm.

Bibliography: p.

ISBN 0-472-10122-6 (alk. paper)

1. Political science—Decision making—Mathematical models.
  2. Elections—Mathematical models.
  3. International relations—Mathematical models.
  4. Legislative bodies—Mathematical models.
- I. Ordeshook, Peter C., 1942-

JA74.M63 1989  
320'.01'51--dc20

89-32107  
CIP

## A Mathematical Proof of Duverger's Law

*Thomas R. Palfrey*

It is not just coincidence that Duverger's Law has established itself as one of the premier empirical regularities in political science. With only minor caveats,<sup>1</sup> it is not only a stylized fact, but also a well-documented fact<sup>2</sup> that single-member-district electoral systems in which winners are decided by simple plurality rule usually produce two-party systems. It is, as far as I know, the only such regularity in political science that is widely referred to as a "law."

What seems surprising, and what motivates this essay, is that in spite of the revolutionary ascent of formal modeling to the forefront of political theory methodology, this clear empirical regularity which has extremely intuitive informal explanations seems to have miraculously escaped the grips of an unambiguous mathematical theorem. If it is as true a law as many seem to believe, then there should be a simple theoretical explanation for it that formalizes the informal stories and rationalizations that have been repeatedly offered for over a century. I hope this would also help identify and illuminate the rare circumstances in which departures from the most sweeping versions of the law might be

---

California Institute of Technology. The author would like to thank the National Science Foundation for financial support. This paper was prepared for delivery at the Economic Theories of Politics conference in Haifa, Israel, June, 1988, and has benefited from discussions with Bruce Cain, John Ledyard, and Richard McKelvey.

1. The most well-known exception is Canada, where provincial party systems are bipartisan, but not all provinces have the same two dominant parties. This produces a patchwork national party system that is a conglomeration of strong regional parties, but in which there are still two dominant national parties. The only contradiction to the law here is that the minor parties do not completely disappear. Apparently, and not surprisingly, parties that have heavily concentrated local support can compete for national office in their regions. It is very important to note, however, that within those regions the local party systems tend to be bipartisan. Furthermore, only the two major parties (the Tories and the Liberals) have controlled the national government. A second exception is India (see Weiner 1957), where there is only one dominant party, but third parties seem to be able to survive. Riker (1976b) argues that this can happen when historical circumstances produce a consensus party that is centrally located on the political spectrum and has wide popular appeal: the Congress Party of India.

2. See Rae (1971) and many of the references in Riker's (1982b) excellent survey of Duverger's Law.

expected. This essay offers such a model that explains Duverger's Law and indicates some unusual circumstances in which the law might not hold up.

What do all the informal explanations have in common? We are fortunate that Riker (1982*b*) has already collected many of these, so I will give only a few representative examples. One of the earliest is presented by Droop (1869) who says that unpopular parties do not receive votes because voters do not want to waste their votes:

Each elector has practically a choice between two candidates or sets of candidates . . . . (T)he electors usually find out that their votes will be thrown away, unless given in favor of one or the other of the parties between whom the election really lies.

This is a remarkable statement that is really a claim about equilibrium behavior of rational voters. It says that voters are making calculating decisions that weigh the chances of affecting the outcome given how other voters vote. It states that strategic voting is rational behavior, that voters are rational and do this, and that it results in a stable equilibrium configuration with only two parties. Thus voters are not only rational in his explanation, but they have rational expectations. His early conjecture about this type of strategic voting has been widely confirmed.<sup>3</sup>

Earlier this century, Schattschneider (1942, 82) stated:

(P)eople who vote for minor opposition parties *waste their votes*. All who oppose the party in power are made to feel a certain need for concentrating their support behind the party most likely to lead a successful opposition. As a consequence, the tendency to support minor parties is checked. The tendency of the single-member district system to give the second major party a great advantage over all minor parties is extremely important. In this way it is possible to explain the *longevity* of the major parties and the instability of the minor parties. (Emphasis as in original)

These two explanations coincide almost exactly with what Duverger (1951, 226) himself called the "psychological factor" leading to two dominant parties:

In cases where there are three parties operating under the simple majority single-ballot system the electors soon realize that their votes are wasted if they continue to give them to the third party: hence their natural tendency

3. See, for example, Cain (1978), Duverger (1950), Bense and Sanders (1979), Spafford (1972), and others.

to transfer their vote to the less evil of its two adversaries in order to prevent the success of the greater evil.

Duverger's psychological factor is then, quite simply, rational strategic voting.

There have been two different formal theoretical approaches which have begun to address the question of what is the stable number of parties in a winner-take-all system. Both approaches are innovative and move a step beyond the standard Downsian models of electoral competition by opening up the possibility of multiparty equilibria.

The first approach captures the aspect of multicandidate competition for a single seat that is the focus of traditional explanations for Duverger's Law. With more than two candidates, voters have a significantly more difficult decision problem, since they may be better off voting for a candidate other than their first choice, if it appears that their first choice stands no realistic chance of victory.<sup>4</sup> This is the direction we pursue here.

The first attempt to formalize this connection between Duverger's Law and rational voting behavior appears in Riker (1976). Ironically, a major point of that essay is to construct a theoretical model that can explain the experience in India with persistent third parties, what Riker (1976, 94) refers to (perhaps overenthusiastically) as "an egregious exception" to Duverger's Law.<sup>5</sup> Cox (1987) adopts a similar approach but embeds the strategic decision making by voters into a probabilistic equilibrium model based on the theory of Bayesian games (Ledyard 1981, 1984). He shows that the only circumstances in which strategic voting might lead voters to abandon their most preferred candidate is if their second most preferred candidate is expected to receive more votes. However, even that is not a sufficient condition for strategic voting. He uses this result to show that there will be a marginal effect of strategic voting to help candidates who are expected to do well and to hurt candidates who are expected to do less well.

This essay obtains a much stronger result. Using the same basic model of voter decision making as Ledyard (1981) and Cox (1987), we show that when the number of voters in the electorate is large, the

4. In fact, this is not only true for single-member district simple-plurality rule systems, but for most other multicandidate rules as well. Cumulative voting, proportional representation, the Borda count, the Hare system, and virtually any other method have this difficulty. For a general statement about the widespread problems of eliciting "sincere" behavior, see Gibbard (1973).

5. As we will show later in the essay, this "egregious exception" is in fact a theoretically predicted exceptional case: the proverbial exception that proves the rule.

equilibrium share of the “third-party vote” must necessarily be small. Moreover, this equilibrium share of the vote declines to zero in the limit as the size of the electorate grows. This result is true under very general assumptions about the heterogeneity and distribution of voter preferences over candidates, and the total number of candidates in the election, and possibilities for abstention.

The necessity of examining large electorates seems reasonable, if one interprets Duverger’s Law as describing a property of national electoral systems. In fact, Riker argues that a *large* electorate should be included as one of the conditions for plurality voting to have such a forceful impact on the party systems (1982*b*, 755). It is this essay’s focus on the asymptotic properties of strategic voting that enables this stronger result to be obtained.

The second of these approaches (Brams and Straffin 1982; Palfrey 1984) is quite different and brings in dynamic, intertemporal factors by investigating whether two established parties will be able to adopt issue positions to preempt the successful entry of any third party. Palfrey (1984) shows in a simple one-dimensional spatial model that an equilibrium configuration of the two established parties in a winner-take-all system (when it exists) will have one leftist party and one party right of center, which are jointly situated so that no third party can enter the competition and win. The intuition is that by positioning themselves neither too close together nor too extreme, the two established parties are located so that there is no “room” for an entrant. Thus, given two established parties, additional parties are not viable.<sup>6</sup> The weakness of this approach is that it assumes no strategic voting.

The rest of the paper is organized as follows. Section II presents the model of three-candidate elections, following the assumptions of Cox (1987). Section III proves the central result of the essay, that in equilibrium, the share of the vote for the third party declines to zero as the number of voters increases. Section IV explores the extent to which these results generalize when candidates are free to select platforms, as opposed to when the distribution of voter preferences over candidates is exogenously fixed.

6. A related model is analyzed by Greenberg and Shepsle (1987). They define an axiomatic notion of stability that is somewhat different from the noncooperative game-theoretic equilibrium used in Palfrey (1984), but that retains the basic idea that established parties are situated in a way that deters entry of new parties. However, they apply their stability notion to election systems with multiple member districts, not winner-take-all systems.

## II. The Model

There are three candidates,  $A, B, C$ . There are  $n$  voters. Each voter has a (strict) preference ranking over the three candidates and we represent a voter’s preferences for each candidate by a Von Neumann–Morgenstern utility number. These utilities are normalized so that voter  $i$  receives a utility of 1 if his (or her) first choice wins, a utility of 0 for his least preferred candidate and a utility of  $v_i$  for his middle-ranked candidate. The distribution of voter preferences is represented by  $\langle q_{AB}, q_{AC}, q_{BA}, q_{BC}, q_{CA}, q_{CB} \rangle$  and  $\langle F_{AB}(\cdot), F_{AC}(\cdot), F_{BA}(\cdot), F_{BC}(\cdot), F_{CA}(\cdot), F_{CB}(\cdot) \rangle$ .

Each  $q_{ij}$  equals the probability a randomly selected voter ranks candidate  $i$  first and  $j$  second, or alternatively, the average frequency of each possible preference ranking in the electorate. We assume each voter’s preference is independently drawn from this distribution. A voter who ranks  $i$  first and  $j$  second is henceforth referred to as an  $ij$ -type. Below, we make two assumptions about the distribution. That these assumptions are relatively innocuous is argued later in the essay.

ASSUMPTION 1.  $q_{ij} > 0$  for all  $i, j = A, B, C$ .

In other words, we assume that every ranking of the candidates can occur (but, possibly, with very low frequency). Each  $F_{ij}(\cdot)$  is the cumulative distribution function (*cdf*) of Von Neumann–Morgenstern utility values of  $j$  (the second-ranked candidate) for a randomly selected  $ij$ -type. An  $ij$ -type who values  $j$  at  $v$  will be called an  $ij$ - $v$ -type.

ASSUMPTION 2. For all  $i, j, F_{ij}(\cdot)$  is twice continuously differentiable, where,

$$\begin{aligned} F_{ij}(0) &= 0 \\ F_{ij}(1) &= 1 \\ f'_{ij}(v) &\equiv F'_{ij}(v) > 0 \text{ for all } v \in [0, 1]. \end{aligned}$$

Assumption 2 rules out mass points of voters and assumes that the probability that two randomly selected voters will have exactly the same ranking and exactly the same “intensity of preference” ( $v$ ) is negligible. Finally, we assume that each voter’s preferences are independently drawn from the probability distribution given

$$P = \langle q_{AB}, \dots, q_{CB}, F_{AB}, \dots, F_{CB} \rangle.$$

Now that we have specified voter preferences, we can specify the voting game. Each voter simultaneously chooses one of the candidates

to vote for, taking as given the voting strategies of other voters, in order to maximize the expected utility of the outcome of the election. Recall that since we are talking about election of a *single* candidate, ties will have to be broken if two (or all three) candidates tie for the most votes received. How these ties are broken does not affect the results in section III, so we will simply assume that ties are broken alphabetically (e.g., *A* beats *B* in a tiebreaker).

Each voter knows only his own preferences between *A*, *B*, and *C* and that the other voters are independently distributed according to *P*. Therefore, a voter views the strategies of the other voters as functions that specify what each other voter would do for every type he might be. We can represent such a strategy for one voter as a measurable function

$$\sigma^i: \{AB, AC, BA, BC, CA, CB\} \times [0, 1] \rightarrow \{A, B, C\}.$$

To simplify the proofs, we will only investigate properties of “symmetric” equilibria. In other words, we will investigate stable behavior in which only two *identical* voters will make identical voting decisions. This restriction enables us to suppress the *i*-index on  $\sigma$ . Therefore, a voter views the strategy of each other voter as the *same* function  $\sigma$ . This symmetry assumption could be dispensed with, but it would be at considerable cost in notation.

The careful reader will have also noticed by now that we assume all voters vote. Thus, in contrast to Ledyard (1981) or Palfrey and Rosenthal (1985) we do not consider the possibility of abstentions.

So far, we have accumulated a number of assumptions that potentially might restrict the scope of the theorem. Summarizing, these are:

1. All preference types are possible.
2. Distribution of preferences has no mass points.
3. Voter preferences are independent.
4. No abstention.
5. Symmetric equilibrium.

We will discuss relaxing these assumptions later in the essay.

Let  $D_A$ ,  $D_B$ , and  $D_C$  denote the set of voter types who vote for *A*, *B*, or *C*, respectively under  $\sigma$ . The probability a randomly selected other voter will vote for *A*, *B*, or *C*, is denoted  $\pi_A$ ,  $\pi_B$ , or  $\pi_C$ , respectively where  $\pi_A + \pi_B + \pi_C = 1$  and  $\pi_j \geq 0 \forall j$ . Then  $\pi_i$  is the probability that a randomly selected voter other than *i* has a type in  $D_j$ . For shorthand, we can denote a strategy  $\sigma = (D_A, D_B, D_C)$ .

Given  $(D_A, D_B, D_C)$  and  $(\pi_A, \pi_B, \pi_C)$ , voter *i* adopts a strategy to

maximize expected utility. This strategy is characterized below for an *AB-v*-type voter. Optimal strategies against  $(D_A, D_B, D_C)$ —“best responses”—for other types of voters are similar.

Finally, we assume no one votes for the candidate ranked last in preference. Thus, an *AB* voter never votes for *C*, etc. This is quite reasonable, as a voter can never gain by doing so. A strategy that calls for voting for one's *least* preferred candidate would be a *weakly dominated strategy*. Ruling out weakly dominated strategies is usually considered reasonable, and the logic of doing so in the present context is compelling.<sup>7</sup>

The following lemma states that any equilibrium strategy must have the property that a voter will vote sincerely (i.e., for his top-ranked candidate) for small values of  $\nu$  (i.e., if his second-ranked and third-ranked candidates both produce relatively low utility levels). A voter will only vote insincerely in equilibrium if  $\nu$  is relatively large (i.e., his second-ranked candidate is sufficiently close in utility to his top-ranked candidate). Furthermore, insincere voting requires that the likelihood of switching the outcome from one's worst candidate to one's second-choice candidate by voting insincerely instead of sincerely exceeds the probability that insincere voting switches the outcome from one's most preferred candidate to one's least preferred candidate.

LEMMA 1. *If other voters use  $(D_A, D_B, D_C)$ , generating probabilities  $(\pi_A, \pi_B, \pi_C)$ , and if *i* is a type *AB-v*-voter, then *i*'s best response is:*

- (i) Vote for *A* if:

$$p_{AB}^n (1 - \nu) + p_{AC}^n > p_{CB}^n \nu$$

- (ii) Either vote for *A* or *B* if (i) holds with equality.

- (iii) Vote for *B* if  $p_{CB}^n \nu > p_{AB}^n (1 - \nu) + p_{AC}^n$ .

where

$$\begin{aligned} p_{AB}^n &= \text{probability that voting for } A \text{ yields } A, \text{ but voting for } B \text{ yields } B \\ p_{AC}^n &= \text{probability that voting for } A \text{ yields } A, \text{ but voting for } B \text{ yields } C \\ p_{CB}^n &= \text{probability that voting for } A \text{ yields } C, \text{ but voting for } B \text{ yields } B. \end{aligned}$$

7. Exceptions to the “reasonableness” of eliminating dominated strategies seem to arise in Prisoners' Dilemmas and related games where there are clear group gains to coordinating behavior. That is, everyone benefits when *everyone* adopts a dominated strategy. Such gains are not present in the voting game considered here. For further discussion of the appropriateness of eliminating dominant strategies see Palfrey and Srivastava (1986).

These probabilities are computed using trinomial formulas, with parameters  $\pi_A$ ,  $\pi_B$ ,  $\pi_C$ , and  $n$ . They are derived below:

$$\begin{aligned} p_{AC}^n &= \text{probability that out of } (n-1) \text{ other voters, } a \text{ vote for } A, \\ &\quad b \text{ vote for } B, c \text{ vote for } C, \text{ and } b < a = c - 1 \\ &= \text{probability}\{b < a = c - 1\} \\ &= \sum_{k=\lfloor \frac{n-1}{3} \rfloor}^{\lfloor \frac{n-1}{2} \rfloor - 1} \frac{(n-1)!}{k!(k+1)!(n-2k-2)!} (\pi_A)^k (\pi_B)^{n-2k-2} (\pi_C)^{k+1} \end{aligned}$$

where  $[X]$  is the least integer greater than or equal to  $X$ . Similarly,

$$\begin{aligned} p_{CB}^n &= \text{probability}\{a < b = c - 1\} \\ &= \sum_{k=\lfloor \frac{n-1}{3} \rfloor}^{\lfloor \frac{n-1}{2} \rfloor - 1} \frac{(n-1)!}{k!(k+1)!(n-2k-2)!} (\pi_A)^{n-2k-2} (\pi_B)^k (\pi_C)^{k+1} \end{aligned}$$

and

$$\begin{aligned} p_{AB}^n &= \text{probability}\{a = b \geq c - 1\} \\ &\quad + \text{probability}\{a = b - 1 \geq c - 1\} \\ &= \sum_{k=\lfloor \frac{n+1}{3} \rfloor - 1}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!(n-2k-1)!} (\pi_A)^k (\pi_B)^k (\pi_C)^{n-2k-1} \\ &\quad + \sum_{k=\lfloor \frac{n+1}{3} \rfloor - 1}^{\lfloor \frac{n-1}{2} \rfloor - 1} \frac{(n-1)!}{k!(k+1)!(n-2k-2)!} (\pi_A)^k (\pi_B)^{k+1} (\pi_C)^{n-2k-2}. \end{aligned}$$

Analogous formulas for the voters with other preference orders over candidates can also be derived. We say  $\sigma$  is a *Symmetric Bayesian-Nash Equilibrium* (or simply *Equilibrium*) if for every  $i$ ,  $\sigma(t_i)$  maximizes  $i$ 's expected utility when  $i$  is type  $t_i$  and when all other voters use the strategy  $\sigma$ .

From Lemma 1, if either  $p_{ij}^n$  or  $p_{kj}^n$  is positive, we may rewrite the inequality in (i) as:

$$v_{ij}^n \geq \frac{p_{ij}^n + p_{ik}^n}{p_{ij}^n + p_{kj}^n} \quad (2)$$

Consequently insincere voting not only requires  $v$  to be relatively large, but also requires that the voter is more likely to be pivotal between his second and last choices than between his first and last choices ( $p_{kj}^n > p_{ik}^n$ ). Therefore, an equilibrium can be written as a set of six *cut points*,  $v_n^* = (v_{AB}^n, v_{AC}^n, v_{BA}^n, v_{BC}^n, v_{CA}^n, v_{CB}^n)$ , each of which must either equal 1 or satisfy (2) with equality. A cut point has a convenient and simple interpretation. A voter votes for his second choice if and only if his intensity of preference for that second choice ( $v$ ) is sufficiently high. Inequality (2) is a precise statement of what "sufficiently high" means. For some  $ij$ -types, there is no  $v \in (0, 1)$  that is sufficiently high, in which case no voter with this order of preference for the candidates will vote strategically. This happens for an  $ij$ -type when  $p_{ik}^n > p_{kj}^n$ .<sup>8</sup>

Lemma 1, then, specifies six inequality conditions that must hold at an equilibrium. However, this is only a partial description of an equilibrium. A further requirement of *Bayesian equilibrium* is that  $\pi_A$ ,  $\pi_B$ , and  $\pi_C$  are in fact generated by the voters all adopting the decision rule given by the equilibrium cut points. This is sometimes referred to as the "rational expectations" property of Bayesian equilibrium, because it means that if all voters share expectations that other voters vote for  $A$ ,  $B$ , and  $C$  with probabilities  $\pi_A$ ,  $\pi_B$ , and  $\pi_C$ , then the optimal behavior by voters, given these expectations, will produce the expected voting pattern—i.e., expectations (or beliefs) are self-fulfilling. This produces three additional equilibria conditions:

$$\begin{aligned} \pi_A &= q_{AB} F_{AB}(v_{AB}^n) + q_{AC} F_{AC}(v_{AC}^n) \\ &\quad + q_{BA} [1 - F_{BA}(v_{BA}^n)] + q_{CA} [1 - F_{CA}(v_{CA}^n)] \end{aligned} \quad (3A)$$

$$\begin{aligned} \pi_B &= q_{BA} F_{BA}(v_{BA}^n) + q_{BC} F_{BC}(v_{BC}^n) \\ &\quad + q_{AB} [1 - F_{AB}(v_{AB}^n)] + q_{CB} [1 - F_{CB}(v_{CB}^n)] \end{aligned} \quad (3B)$$

$$\begin{aligned} \pi_C &= q_{CA} F_{CA}(v_{CA}^n) + q_{CB} F_{CB}(v_{CB}^n) \\ &\quad + q_{AC} [1 - F_{AC}(v_{AC}^n)] + q_{BC} [1 - F_{BC}(v_{BC}^n)] \end{aligned} \quad (3C)$$

#### Possible Equilibria

We distinguish between three possible types of equilibria, depending upon the configuration of voter strategies. These are called *one-party equilibrium*, *two-party equilibrium*, and *three-party equilibrium*, and are defined by the number of parties whose candidate receives a positive expected vote. The form of Duverger's Law that we intend to prove is that, within the confines of our model, only two-party equilibrium occur.

8. Cox (1987) has pointed out that one case where this happens is when  $\pi_i > \pi_r$ .

Several observations can now be made about equilibrium with at least four voters. First, there will exist an equilibrium. In fact there exists *at least three equilibria* where in each of these equilibria there is substantial strategic voting, in the sense that some voters do not vote for their first choice. To see this, take any pair of parties, assume all voters vote for their preferred one of these two, and consider the implications of Lemma 1: A voter should vote for his first choice unless his first choice is not one of the two candidates receiving votes, in which case he should vote for his second choice. (This argument requires that the population of voters be at least four.) This is an equilibrium.

Next, consider a possible one-party equilibrium in which the candidate of one of the parties receives all of the votes. Why is this not an equilibrium? The probability of casting a decisive vote is 0, so each voter is indifferent between voting for any candidate, since his vote cannot affect the outcome (assuming the population of voters exceeds two). However, for a voter who ranks the unanimous winner *last*, voting for either other candidate is guaranteed to generate at least as high an expected utility as voting for the candidate who is expected to get all the votes, and, for some (out of equilibrium) configuration of strategies by other voters, generates strictly higher expected utility. Thus such equilibria rely on the use of dominated strategies. This eliminates the “unanimous” equilibria.

To summarize, Lemma 1 and the assumption that voters do not use dominated strategies together immediately imply that:

1. There is no equilibrium in which only one party's candidate receives votes.
2. There are three two-party equilibria.

We next turn to the central question of this paper: In large electorates, (as  $n \rightarrow \infty$ ) will there ever exist any three-party equilibrium? The answer is no, except for knife-edge cases.

### III. Strategic Voting with Many Voters

The objective of this section is to show that in large electorates with simple-plurality, single-ballot elections there do not exist any equilibria in which all parties are viable. In particular, except for knife-edge cases, the only equilibria that can exist in large electorates involve exactly two parties with candidates who receive a positive fraction of the vote, if we assume voters do not use dominated strategies (i.e., they never vote for their *last* choice).

With three candidates and *any* number of voters, the critical element in the voter calculus is the probability of casting a decisive vote, and in particular, the relative probabilities of casting a decisive vote between one's first and last choices versus one's second and last choices. Consequently, the key aspect of “proving” Duverger's Law in large electorates involves identifying the asymptotic properties of these probabilities. The following lemma turns out to be very important in establishing these properties:

LEMMA 2. Assume  $\pi_A < \min \{\pi_B, \pi_C\}$ . Then  $\lim_{n \rightarrow \infty} \left( \frac{p_{AC}^n}{p_{CB}^n} \right) = 0$ .

*Proof.* The proof consists of a messy formalization of the following intuitive argument. Without loss of generality, assume  $\pi_B \leq \pi_C$ . Then  $p_{AC}^n$  is the probability that the least popular candidate and the “favorite” are tied for the most votes and  $p_{CB}^n$  is the probability that the second most popular and the favorite are tied for the most votes.<sup>9</sup> Thus  $p_{AC}^n / p_{CB}^n$  is equivalent to the ratio of likelihoods that the least popular candidate is tied for victory with the favorite compared to the second most popular candidate being tied with the favorite, *conditional* on one of these two ties occurring. The lemma states that *with many voters, if some candidate ties the favorite it will almost certainly have to be the second most popular candidate*. The implication of this is that if your *personal* favorite is the least popular candidate then, regardless of your relative valuation,  $v$ , of your second choice, with enough voters in the electorate you will be better off voting for your second choice.

#### Outline of Formal Proof (Details in Appendix)

For each  $n$ , denote by  $E_{AC}^n$  and  $E_{CB}^n$  the expected winning share of the two candidates who receive the most votes, conditional on  $A$  and  $C$  being tied for the most, or  $B$  and  $C$  being tied for the most,<sup>10</sup> respectively, and define  $E_{AC} = \lim_{n \rightarrow \infty} E_{AC}^n$  and  $E_{CB} = \lim_{n \rightarrow \infty} E_{CB}^n$ . The proof next proceeds through a series of steps:

9. Actually, because of the tie-breaking rule,  $P_{AC}^n$  equals the probability that, with  $n - 1$  voters,  $C$  receives the most votes and  $A$  receives *one less* vote than  $C$ ;  $P_{CB}^n$  equals the probability that  $C$  receives the most votes and  $B$  receives one less than  $C$ ;  $P_{AB}^n$  equals the probability  $A$  and  $B$  tie for the most votes or  $B$  gets one more vote than  $A$ .

10. See note 9.

Step 1: Prove that

$$E_{AC} = \begin{cases} \frac{1}{3} & \text{if } (\pi_B)^2 > \pi_A \pi_C \\ \frac{1}{2 + \frac{\pi_B}{\sqrt{\pi_A \pi_C}}} & \text{if } (\pi_B)^2 < \pi_A \pi_C \end{cases}$$

$$E_{CB} = \frac{1}{2 + \frac{\pi_A}{\sqrt{\pi_B \pi_C}}}$$

Step 2: Let

$$\bar{p}_{AC}^n = \frac{(n-1)! (\pi_A \pi_C)^{(n-1)E_{AC}} \pi_C \pi_B^{n-2-2E_{AC}(n-1)}}{[(n-1)E_{AC}]! [1 + (n-1)E_{AC}]! [n-2-2E_{AC}(n-1)]!}$$

$$\bar{p}_{CB}^n = \frac{(n-1)! (\pi_B \pi_C)^{(n-1)E_{CB}} \pi_C \pi_A^{n-2-2E_{CB}(n-1)}}{[(n-1)E_{CB}]! [1 + (n-1)E_{CB}]! [n-2-2E_{CB}(n-1)]!}$$

Prove:

$$\lim_{n \rightarrow \infty} \left( \frac{\bar{p}_{AC}^n}{\bar{p}_{CB}^n} \right) = 0$$

Step 3:

$$\lim_{n \rightarrow \infty} \left( \frac{\bar{p}_{AC}^n}{\bar{p}_{CB}^n} \right) = 0 \Leftrightarrow \lim_{n \rightarrow \infty} \left( \frac{p_{AC}^n}{p_{CB}^n} \right) = 0$$

These three steps are proved in the appendix and establish the lemma in case  $\pi_C \geq \pi_B$ . If  $\pi_B \geq \pi_C$ , the entire proof goes through identically with the only change being that  $E_{AC} = 1/3$  for all values of  $(\pi_A, \pi_B, \pi_C)$ , such that  $\pi_B \geq \pi_C$ .

**THEOREM.** Fix  $F$ . There does not exist a sequence of equilibria  $\{v_n^*\}_{n=1}^\infty$  such that  $0 < \pi_A^* < \min\{\pi_B^*, \pi_C^*\}$ , where  $\pi_j^* = \lim_{n \rightarrow \infty} \pi_j(v_n^*)$ ,  $j = A, B, C$ , where  $\pi_j(v_n^*)$  is given by equation (3j).

*Proof.* Without loss of generality, let  $\pi_B^* = \min\{\pi_B^*, \pi_C^*\}$  and hypothesize  $0 < \pi_A^* < \pi_B^*$ . Then for some profile, there must exist some subsequence of  $\{v_n^*\}$ , call it  $\{v_m^*\}$ , such that at least one of the following limits holds:

$$\lim_{m \rightarrow \infty} v_{AB}^{*m} = v_{AB}^* > 0$$

$$\lim_{m \rightarrow \infty} v_{AC}^{*m} = v_{AC}^* > 0$$

$$\lim_{m \rightarrow \infty} v_{BA}^{*m} = v_{BA}^* < 1$$

$$\lim_{m \rightarrow \infty} v_{CA}^{*m} = v_{CA}^* < 1.$$

In other words, there must be some region with a positive measure of voters who vote for  $A$  in the limit. We will now apply Lemma 1 to prove that  $v_{AB}^*$  and  $v_{AC}^*$  must equal 0 and  $v_{BA}^*$  and  $v_{CA}^*$  must equal 1. From Equation (2), we have

$$v_{AB}^{*m} = \frac{p_{AB}^m + p_{AC}^m}{p_{AB}^m + p_{CB}^m} = \frac{p_{AB}^m/p_{CB}^m + p_{AC}^m/p_{CB}^m}{p_{AB}^m/p_{CB}^m + 1}$$

(Note that for large  $m$   $p_{CB}^m > 0$ , by the hypothesis that  $\pi_B^* > 0$  and  $\pi_C^* > 0$ .) If limits exist, then

$$v_{AB}^* = \lim_{m \rightarrow \infty} v_{AB}^{*m} = \frac{\lim_{m \rightarrow \infty} (p_{AB}^m/p_{CB}^m) + \lim_{m \rightarrow \infty} (p_{AC}^m/p_{CB}^m)}{\lim_{m \rightarrow \infty} (p_{AB}^m/p_{CB}^m) + 1}$$

From Lemma 1, each of the limits on the right-hand side of the equation above exists and equals 0, so  $v_{AB}^* = 0$ . By the same argument,  $v_{AC}^* = 0$ . That  $v_{AB}^* = 1$  follows simply from the observation that if every voter who prefers  $A$  to  $B$  (and  $C$  is last) will never vote for  $A$ , then no voter who prefers  $B$  to  $A$  (and  $C$  is last) will ever vote for  $A$ . Similarly  $v_{CA}^* = 1$ . Therefore,  $\pi_A^* = 0$ , a contradiction. (Q.E.D.)

An alternative statement of the proof is the following.

**THEOREM.** Fix  $F$  and fix  $\pi_A > \pi_B > \pi_C > 0$ . Then there is an electorate size  $N$  such that for all  $n \geq N$  and for all equilibria  $\{v_n^*\}$ ,  $(\pi_A^*(v_n^*), \pi_B^*(v_n^*), \pi_C^*(v_n^*)) \neq (\pi_A, \pi_B, \pi_C)$ , and  $\pi_C^*(v_n^*) < \pi_C$ .

In other words, the least popular party's equilibrium share of the vote must converge to 0 as the electorate becomes large, if all voters are acting strategically.

**Exceptional Cases**

The above theorem does not completely rule out three-party equilibria since there is the possibility that  $\pi_A = \min \{\pi_B, \pi_C\}$ . There are two cases to consider here. First, if  $\pi_B = \pi_C$ , then we must have  $\pi_A = \pi_B = \pi_C = 1/3$ . In this case, for large  $n$ , all voters vote for their first choice, since the right-hand side of inequality Equation (2) converges to 1. This can be sustained as an equilibrium only if  $q_{AB} + q_{AC} = q_{BA} + q_{BC} = q_{CA} + q_{CB}$ , a knife-edge case.

Second, if  $\pi_B = \pi_C$ , then  $\pi_A > \pi_B = \pi_C$ . In this case, inspection of inequality Equation (2) for large  $n$  shows that in equilibrium we must have all voters voting for their first choice, so such an equilibrium can occur in large electorates only if either  $q_{AB} + q_{AC} = q_{BA} + q_{BC}$  or  $q_{AB} + q_{AC} = q_{CA} + q_{CB}$  or  $q_{BA} + q_{BC} = q_{CA} + q_{CB}$ . Again, these are exceptional cases.

The exceptional cases have interesting interpretations. The first case, where  $\pi_A = \pi_B = \pi_C = 1/3$  can occur in one of two ways. The obvious way it can occur is ruled out by assumption and might arise if all voters evaluate all candidates identically. While this may not seem to be a likely thing to happen from an empirical standpoint, and it may seem intuitively implausible, it certainly cannot be ruled out theoretically; in fact some recent versions of the Downsian model of candidate competition make such a theoretical prediction. We return to this in the next section.

This case can also arise if the parties take positions so that voters are equally split in their preferences between the parties. While there maybe some rare historical instances where this has happened, such situations seem inherently unstable because of the delicate balance on which *exactly* equal three-way division hangs.

The second exceptional case arises when there is one very popular party and two (or more) equally minor parties. Again I would argue that is relatively unlikely to happen, but here we do have a specific empirical case that seems to fit: India. The discussion in Riker (1976b) and some of the references he cites support this theoretical explanation quite convincingly. The Congress party is clearly dominant relative to a collection of lesser parties, the strongest of which are roughly equal in strength. The lesser parties are of significant size and do not disappear, but are just too disparate to unify into a single, effective opposition party.

**IV. Extensions and Generalizations**

**More Than Three Possible Parties**

The basic idea of the results in the last section was that if exactly three parties put up candidates for election when there are many voters, then equilibrium voting patterns will end up eliminating *exactly one* party in

the sense that its candidate must receive an order of magnitude fewer votes than the runner-up. One (of several) artificial features of the model was that only three parties were considered. Suppose  $m > 3$  parties competed. Then is it possible that perhaps more than 2, say  $m - 1$ , parties could be viable in equilibria? The answer is no. *Two* is unequivocally special in this regard. Suppose we wanted to see if the expected vote shares,  $\pi_A \geq \pi_B > \pi_C \geq \pi_D \geq \dots > 0$ , in an  $m$ -party race could possibly be a limit point of equilibrium expected vote shares as the number of voters became large. As before, we find (by a similar argument) that the probability of a voter being decisive between  $A$  and  $B$  will become (in the limit) infinitely greater than the probability of a voter being decisive between  $A$  and any other pair. Thus, in the natural generalization of this model to plurality voting between candidates from  $m$  parties, Duverger's Law remains intact.

It is probably instructive to sketch the proof. What has to be shown is that in large electorates, equilibrium voting behavior implies that a voter will always vote for the most preferred candidate of the two frontrunners. To see this, consider a voter with utility  $v_1$  for his or her first choice,  $v_2$  for second, etc., up to  $v_m$  for his least-preferred candidate. That is  $v_1 \geq v_2 \geq \dots \geq v_m$ . Let  $v_k$  denote the utility of his first choice among the frontrunners (i.e.,  $k$  equals either  $A$  or  $B$ ) and let  $v_l$  denote the utility of his second choice among the frontrunners, and assume  $v_k > v_l$ .<sup>11</sup> Let  $j$  be any candidate other than  $k$ . We can write down a condition analogous to inequality Equation (2) that specifies when such a voter is better off voting for  $k$  than  $j$ . Using the same notation as before, we get for given  $n$ :

$$v_k > v_j \left[ \frac{\sum_{i \neq j} p_{ij}^n}{\sum_{h \neq k} p_{kh}^n} \right] + \sum_{i \neq j, k} v_i \left[ \frac{p_{ki}^n - p_{ij}^n}{\sum_{h \neq k} p_{kh}^n} \right] \tag{2'}$$

This can be rewritten as:

$$v_k > v_j \left[ \frac{\sum_{i \neq j} \left( \frac{p_{ij}^n}{p_{kl}^n} \right)}{\sum_{h \neq k} \left( \frac{p_{kh}^n}{p_{kl}^n} \right)} \right] + \sum_{i \neq j, k} v_i \left[ \frac{(p_{ki}^n - p_{ij}^n)}{p_{kl}^n} \right] \tag{2''}$$

11. By assumption  $v_k > v_l$  for all except a measure zero set of voters.

The limiting argument now proceeds in the familiar fashion. Fixing  $\pi_k \geq \pi_l > \dots > 0$ , we get  $p_{ij}^n/p_{kl}^n \rightarrow 0$ , for all  $ij \neq kl$ . Thus, the right-hand side of Equation (2'') converges to  $v_j$ , independently of  $j$ .

#### Spatial Party Competition

A natural question to ask is whether the model presented here is consistent with a spatial model of voter behavior. The answer is yes, but with a caveat. If the voters are drawn from a continuous distribution of "reasonable" (say, Euclidean) preferences and parties offer candidates at different points in the policy space then the results go through directly. If the parties generate candidates over which all voters are indifferent the result will be false. Therefore, the caveat is that the theorem applies only to differentiated parties.

Several remarks about this implicit assumption that three candidates do not converge are in order. First, even if parties do converge in policy space, the results will still go through as long as voters systematically misperceive the locations of the candidates, or if candidates have nonpolicy attributes that distinguish them (age, looks, voice quality, ethnic background, regional base, etc.). The failure of the result to hold *at convergence of candidates* is appropriately viewed as a knife-edge case.

Second, this knife-edge feature suggests that the next step to take is to embed a game of party competition within the structure of the  $n$ -party voter equilibrium studied here. This would generate a model somewhat like Ledyard (1984), with two important exceptions. First, we have three candidates, not two. Second, Ledyard allows for abstention. While the abstentions are easy to incorporate, the complications generated by more than two candidates are subtle and difficult because of the inevitability of strategic voting. In Ledyard's (1984) model,<sup>12</sup> the formal structure is a two-stage game. The first stage of the game is a simultaneous choice of platforms (or nomination of candidates) by each party. In the second stage, the voters get to observe (perfectly) these choices by candidates but have incomplete information about the preferences of the other voters. Voters then choose simultaneously which candidate to vote for. The second-stage subgames correspond exactly with the analysis in the body of this essay. However, as we pointed out earlier, there are multiple voter equilibria in *any* subgame (i.e., for any location of the candidates). These multiple equilibria make the equilibrium behavior in

12. This is also the case in other models in which candidates and noncandidates are both treated as strategic actors (see, for example, Austen-Smith 1987a and Ingberman 1986, in the context of campaign financing).

the first stage of the game essentially indeterminate—i.e., there will be many candidate equilibria that can be supported by some configuration of voter equilibria in the subgames. (By *configuration* we mean a function that assigns party platforms into equilibrium voting strategies.) This indeterminacy needs to be investigated more thoroughly, but such an investigation is clearly beyond the scope of this essay.

The spatial interpretation of the model presented in section II also relates to Assumption 1, that all preference orders have positive probability. In one dimension with single-peaked preferences, some preferences are impossible. Nonetheless, the theorem goes through even if  $q_{jk} = 0$  for some orderings. Similarly, with respect to the details of Assumption 2, it is not necessary that the density function be positive for all  $v \in [0,1]$ . The only important part of Assumption 2 is the requirement that there be no mass points.

#### Abstention

Any model of voting that explores the implications of rational strategic behavior in *large electorates* is obliged to consider the implications of abstention. It is established by Palfrey and Rosenthal (1985) using a very similar Bayesian game approach that in large electorates rational voters with positive net costs of voting will not turn out. This means, simply, that we must restrict analysis to the portion of the electorate who have nonpositive net costs of voting. There are many reasons why net costs may not be positive: consumption value from voting, feelings of citizen duty or obligation, fear of reprisals or sanctions for not voting (badgering by spouse, union official, friends, party activists), to name a few that are often suggested. In fact, as long as any strictly positive fraction of the electorate has nonpositive net costs of voting, and therefore vote, all of the results of the essay hold.

In an extreme case, if all voters were to have strictly positive net voting costs then the prediction would be that nearly 0 percent turnout would occur with large elections. While this will undoubtedly change the theorems in this essay (I have not worked out the details), the implication of no turnout is wildly out of sync with the significant numbers we observe voting.

#### Dynamics

The results in this essay have been interpreted as implying that if three parties are competing in a single-ballot winner-take-all system, eventually (i.e., in an equilibrium) one of the parties will be weeded out. However, the model is static, even though the interpretation has dynamic overtones. A dynamic model would probably serve two very use-

ful purposes here. First, the equilibrium expectations of voters, which drive their equilibrium voting decisions in multicandidate elections do not emerge from thin air. These expectations are the product of past electoral outcomes, polls, media commentary, and other factors. Polls, past outcomes, etc. themselves are endogenously a function of voters' expectations and decisions. This suggests that it would make more sense from a standpoint of realism to explicitly model expectations and strategies as evolving over time. (For examples of this approach in dynamic economic models, see Marcet and Sargent 1985 and Woodford 1986.)

The second potential contribution of a dynamic model would be to lead toward a resolution of the indeterminacy of multiple equilibria in the static model. The static model in this essay went a long way in this regard, by essentially eliminating all equilibria in the voting game except for two-party equilibria. However, there are still (at least three) multiple two-party voter equilibria, and these in turn produce a plethora of equilibria in the larger game of party competition, as argued above. A well-formulated model of the dynamics of the formation of expectations and beliefs about when a party is a "viable" one (worth expending one's vote on) might shed some light on the indeterminacy issue.

#### Other Electoral Systems

The original statement of Duverger's Law actually has more components than the one examined here. In fact, Duverger asserts a partial converse (which he does not refer to as a "law") of the result proved here, that "the single-majority systems with *second ballot* and *proportional representation* favor multi-partisan" (1951, 239). This presents some new problems from a theoretical standpoint, as the modeling of proportional representation systems is much harder because of more of a need to explicitly study the coalition formation process. This is not to say that in simple majority, single-ballot, non-PR systems, the coalition formation process among legislators is inconsequential. In fact, this is an aspect of the problem we have glossed over in this essay. Usually a party runs a slate of candidates, often across a range of political "districts" to produce a legislature as a concoction of regional winners. The representative system modeled in this paper consisted of but a single "at large" district with a single representative.

Duverger claims that an additional (beside strategic voting) reinforcing feature in preventing more than two parties is the underrepresentation of small parties (share of seats much less than share of votes) that results as a consequence of single-ballot winner-take-all elections at the district level. Since this "mechanical" (Duverger 1951, 226) factor would seem to reinforce the strategic voting factor analyzed in this essay, its

absence from the model does not seem to detract significantly from the results.

On the other hand, the entire notion of PR makes little sense at all in the context of electing a one-member government. The fact that a party can win some seats with less than a majority only makes sense if there are several seats up for grabs. Thus any model of PR will require a model of multimember assemblies. This is a more difficult issue. There are many variations on PR and majority-based electoral systems, as well as completely different types of systems (approval voting, cumulative voting, etc.) for which the implications of strategic voting are not well understood.

#### APPENDIX

*Proof of Step 1 of Lemma 2.* This problem is equivalent to the following one. A total of  $n$  balls are drawn with replacement from an urn containing a large number of red balls, white balls, and green balls that occur with frequencies  $0 < p_1 < p_2 \leq p_3$ , respectively. Let  $X_n^{13}$  denote the random variable taking on values between  $\frac{1}{3}$  and  $\frac{2}{3}$  equal to the proportion of red balls conditional on exactly  $k$  red balls being drawn, exactly  $(k + 1)$  green balls being drawn, and exactly  $(n - 2k - 2)$  white balls being drawn from a sample of  $n - 1$  balls and  $k \geq (n - 1)/3$ . Then

$$E_{13}^n = \frac{1}{n} \sum_{k=\lfloor \frac{n-1}{3} \rfloor}^{\lfloor \frac{n}{3} \rfloor - 1} k \bar{p}_{13}^{kn}, \text{ where } \bar{p}_{13}^{kn} = \text{prob}(X_n^{13} = k/n - 1).$$

Think of the red balls as being votes for  $A$ , the green balls as being votes for  $C$ , and the white balls as being votes for  $B$ .

Standard results in probability theory tell us that the distribution of  $X_n^{13}$  is asymptotically normal. Therefore, since in our problem  $\bar{p}_{13}^{kn}$  is unimodal in  $k$  for each  $n$  and that mode converges to a limit, then the limit of the modes must equal  $E_{13}$ .

The limit of the modes is simple to compute because  $X_n^{13}$  has a distribution similar to a binomial. In particular,

$$\bar{p}_{13}^{kn} = \frac{p_{13}^{kn}}{\sum_{j=\lfloor \frac{n-1}{3} \rfloor}^{\lfloor \frac{n}{3} \rfloor - 1} p_{13}^{jn}}$$

where

$$p_{13}^{kn} = \frac{(n-1)!}{j!(j+1)!(n-2j-2)!} p_1^j p_2^{n-2j-2} p_3^{j+1},$$

$$j = \left\lfloor \frac{n-1}{3} \right\rfloor, \dots, \left\lfloor \frac{n}{2} \right\rfloor - 1.$$

Therefore,  $\bar{p}_{13}^{kn}$  is unimodal in  $k$  if and only if  $p_{13}^{kn}$  is unimodal in  $k$ . To see that  $p_{13}^{kn}$  is unimodal look at first differences,  $\Delta p_{13}^{kn} = p_{13}^{kn} - p_{13}^{(k-1)n}$ . This gives

$$\Delta p_{13}^{kn} = \left[ \frac{(n-2k)(n-2k-1)}{k(k+1)} \left( \frac{p_1 p_3}{p_2^2} \right) - 1 \right] \frac{(n-1)!}{(k-1)!k!(n-2k)!} p_1^{k-1} p_2^{n-2k} p_3^k.$$

So  $\Delta p_{13}^{kn} > 0$  if

$$\frac{p_1 p_3}{(p_2)^2} > \frac{k(k+1)}{(n-2k)(n-2k-1)}.$$

Rewriting this expression using the notation of the lemma, we get  $p_{AC}^{kn}$  is increasing if

$$\frac{\pi_A \pi_C}{(\pi_B)^2} > \frac{k(k+1)}{(n-2k)(n-2k-1)}.$$

The right-hand side is strictly increasing in  $k$ , so  $\bar{p}_{AC}^{kn}$  is unimodal in  $k$ , as desired. If the mode is interior to  $(\frac{1}{3}, \frac{1}{2})$ , then for large  $n$ , it occurs at  $k$  which approximates

$$\frac{\pi_A \pi_C}{\pi_B^2} = \frac{k^2}{(n-2k)^2}$$

or

$$\frac{\pi_A \pi_C}{\pi_B^2} = \frac{(k/n)^2}{[1-2(k/n)]^2}.$$

Solving, gives

$$\frac{k}{n} = \frac{1}{2 + \pi_B / \sqrt{\pi_A \pi_C}}$$

if  $\pi_B^2 > \pi_A \pi_C$ .

Since  $k/n$  is constrained to be at least  $\frac{1}{3}$ , we must have

$$\frac{k}{n} = \frac{1}{3}$$

for  $\pi_B^2 > \pi_A \pi_C$ . Consequently, we get

$$E_{AC} = \begin{cases} \frac{1}{3} & \text{if } \pi_B^2 > \pi_A \pi_C \\ \frac{1}{2 + \pi_B / \sqrt{\pi_A \pi_C}} & \text{if } \pi_B^2 < \pi_A \pi_C \end{cases}.$$

An identical argument proves that  $E_{CB} = \frac{1}{2 + \pi_A / \sqrt{\pi_B \pi_C}}$  because  $\pi_A^2 < \pi_B \pi_C$  by hypothesis. (Recall  $\pi_A < \pi_B \leq \pi_C$ .)

*Proof of Step 2 of Lemma 2:* Let

$$\bar{p}_{CB}^n = \frac{(n-1)! (\pi_B \pi_C)^{(n-1)E_{AC}} \pi_C \pi_A^{n-2-2(n-1)E_{AC}}}{[(n-1)E_{AC}]! (1 + [(n-1)E_{AC}])! (n-2-2[(n-1)E_{AC}])!}.$$

Then

$$\frac{\bar{p}_{AC}^n}{\bar{p}_{CB}^n} = \left[ \frac{\pi_A}{\pi_B} \right]^{1+3[E_{AC}(n-1)]-(n-1)}$$

which, for large  $n$  is bounded above by 1. By construction,  $\bar{p}_{CB}^n < \bar{p}_{CB}^n$ , and it is easy to show that

$$\frac{\bar{p}_{CB}^n}{\bar{p}_{CB}^n} \text{ converges to } 0,$$

so:

$$\lim_{n \rightarrow \infty} \frac{\bar{p}_{AC}^n}{\bar{p}_{CB}^n} = \lim_{n \rightarrow \infty} \left( \frac{\bar{p}_{AC}^n}{\bar{p}_{CB}^n} \right) \left( \frac{\bar{p}_{CB}^n}{\bar{p}_{CB}^n} \right) = 0.$$

*Proof of Step 3 of Lemma 2.* The central limit theorem insures that for any  $\varepsilon > 0$ ,  $p_{AC}^n(\varepsilon)$  and  $p_{CB}^n(\varepsilon)$  converge to  $p_{AC}^n$  and  $p_{CB}^n$  in the sense that  $\lim_{n \rightarrow \infty} (p_{AC}^n(\varepsilon)/p_{AC}^n) = 1$ , and  $\lim_{n \rightarrow \infty} (p_{CB}^n(\varepsilon)/p_{CB}^n) = 1$ , where

$$p_{AC}^n(\varepsilon) = \frac{\min\{\frac{n-1}{2}, \{(n-1)(E_{AC} + \varepsilon)\}\}}{\sum_{k=\max\{\frac{n-1}{3}, \{(n-1)(E_{AC} - \varepsilon)\}\}}^{\{(n-1)(E_{AC} + \varepsilon)\}} \frac{(n-1)!}{k!(k+1)!(2n-2k-2)!} \pi_A^k \pi_B^{n-2k-2} \pi_C^{k-1}}$$

$$p_{CB}^n(\varepsilon) = \frac{\min\{\frac{n-1}{2}, \{(n-1)(E_{CB} + \varepsilon)\}\}}{\sum_{k=\max\{\frac{n-1}{3}, \{(n-1)(E_{CB} - \varepsilon)\}\}}^{\{(n-1)(E_{CB} + \varepsilon)\}} \frac{(n-1)!}{k!(k+1)!(n-2k-2)!} \pi_A^{n-2k-2} \pi_B^k \pi_C^{k-1}.$$

Since  $E_{AC}$  and  $E_{CB}$  are both strictly less than  $1/2$ , we may choose  $\varepsilon$  so that

$$\left(\frac{n-1}{2}\right) - \{(n-1)(E_{AC} + \varepsilon)\} \quad \text{and} \quad \left(\frac{n-1}{2}\right) - \{(n-1)(E_{CB} + \varepsilon)\}$$

both diverge to  $\infty$  in the limit. Since  $E_{CB}$  is strictly greater than  $1/3$ ,  $\varepsilon$  can also be chosen so that

$$\{(n-1)(E_{CB} - \varepsilon)\} - \frac{n-1}{3}$$

diverges to  $\infty$  in the limit. Therefore, we may rewrite  $p_{AC}^n(\varepsilon)$  and  $p_{CB}^n(\varepsilon)$  for  $n$  sufficiently large and  $\varepsilon$  sufficiently small as:

$$p_{AC}^n(\varepsilon) = \frac{\sum_{k=\{(n-1)(E_{AC} - \varepsilon)\}}^{\{(n-1)(E_{AC} + \varepsilon)\}} \frac{(n-1)!}{k!(k+1)!(n-2k-2)!} (\pi_A^k \pi_B^{n-2k-2} \pi_C^{k+1} \delta_k)}{\sum_{k=\{(n-1)(E_{AC} - \varepsilon)\}}^{\{(n-1)(E_{AC} + \varepsilon)\}} \frac{(n-1)!}{k!(k+1)!(n-2k-2)!} (\pi_A^k \pi_B^{n-2k-2} \pi_C^{k+1} \delta_k)}$$

$$p_{CB}^n(\varepsilon) = \frac{\sum_{k=\{(n-1)(E_{CB} - \varepsilon)\}}^{\{(n-1)(E_{CB} + \varepsilon)\}} \frac{(n-1)!}{k!(k+1)!(n-2k-2)!} (\pi_A^{n-2k-2} \pi_B^k \pi_C^{k+1})}{\sum_{k=\{(n-1)(E_{CB} - \varepsilon)\}}^{\{(n-1)(E_{CB} + \varepsilon)\}} \frac{(n-1)!}{k!(k+1)!(n-2k-2)!} (\pi_A^{n-2k-2} \pi_B^k \pi_C^{k+1})}$$

where

$$\delta_k = 0 \text{ if } k < \frac{n-1}{3} \\ = 1 \text{ if } k \geq \frac{n-1}{3}.$$

For each  $n$  both of these sums contain equal numbers of terms that can be matched up. For given  $n$ , call the terms on the right-hand side of the  $p_{AC}^n(\varepsilon)$  equation  $a_1, \dots, a_{m_n}$  and the terms on the right-hand side of the  $p_{CB}^n(\varepsilon)$  equation  $b_1, \dots, b_{m_n}$ . By construction, we have  $\lim a_i/b_i = 0$  uniformly in  $l$ . We also have

$$\lim_{n \rightarrow \infty} \frac{p_{AC}^n(\varepsilon)}{p_{CB}^n(\varepsilon)} = \lim_{n \rightarrow \infty} \frac{a_1 + \dots + a_{m_n}}{b_1 + \dots + b_{m_n}}$$

but

$$\frac{a_1 + \dots + a_{m_n}}{b_1 + \dots + b_{m_n}} \leq \max_{1 \leq i \leq m_n} \frac{a_i}{b_i}.$$

Since  $\frac{a_i}{b_i} \rightarrow 0$  uniformly, we get

$$\lim_{n \rightarrow \infty} \frac{p_{AC}^n(\varepsilon)}{p_{CB}^n(\varepsilon)} = 0.$$

Therefore

$$\lim_{n \rightarrow \infty} \frac{p_{AC}^n}{p_{CB}^n} = 0.$$