Voting and Electoral Competition

Thesis by

Steven Callander

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2002
(Submitted May 30, 2001)

Dedication

To my Parents,

for providing me with opportunities they never had.

Acknowledgments

This thesis has benefited from the guidance of two advisors. It was begun under Jeff Banks who, sadly, passed away before its completion. I thank him for the valuable time he so generously granted me. His impact on the way I work will be with me long after this thesis is done. Richard McKelvey generously and unhesitantly stepped forward during the middle of my job market and took me through the rest of the way. Though I wish circumstances hadn't given me this opportunity to work closely with him, I have enjoyed and benefited from his unique insights. I also appreciate the endless advice and amusement provided by all faculty and graduate students in the Social Sciences that has made this a wonderful time in my life. Of course, I also thank Caltech for putting on the whole show. And, because she is always fighting for the final word, I finish by thanking Catherine Wilson for having the devotion to tolerate me and the many nuances of Australian slang.

Abstract

The behavior of individuals and groups in the political realm is subject to many and varied incentives. These incentives impact significantly not only the candidates who win elections, but also the policies that they implement. This thesis analyzes several aspects of this problem that have until now gone unexplained.

Part 1 contains two models of candidate competition. Chapter 1 details a model of competition under the plurality rule that simultaneously explains two well-documented empirical regularities: that typically only two parties compete in each election (Duverger's Law), and that these parties choose non-centrist policy platforms. I show that if, and only if, competition is for multiple districts does an equilibrium consistent with these phenomena exist. I characterize bounds on district heterogeneity for this to be true, which can be interpreted as describing the domain for Duverger's Law. In Chapter 2, I turn attention to the run-off rule and study a similar model to that of Chapter 1. I find that this subtle change to the counting rule has a significant impact on the incentives and equilibria of the model. In the traditional single district environment there now exists a continuum of two-party non-centrist equilibria, which are robust to simultaneous competition for multiple districts.

In Part 2, I investigate the behavior of voters, and particularly the effect of vote timing on voter behavior and election outcomes. In Chapter 3, I study a model of sequential voting and explain when and why the commonly observed phenomena of bandwagons and momentum arise. I show that only if voters have a desire to vote for the winning candidate, in addition to their desire to select the better candidate, is momentum observed and bandwagons begun. In Chapter 4, I compare these results with analogous results for when voting is simultaneous and characterize when each process is superior. The conclusions confirm commonly held views about the front-loading of U.S. presidential primaries: that in tight races a simultaneous vote is preferred, but in lopsided races a sequential vote is better. Strangely, the superior

performance of sequential voting in lopsided races is precisely because bandwagons occur.

Contents

List of Figures

List of Tables

0.1 Introduction

The formal study of politics, like all social sciences, is ultimately interested in outcomes. The end goal is to understand how different electoral systems and democratic institutions affect the policies that are implemented and the welfare that results. Such an understanding requires careful study of the participants in these processes and the institutions in which they interact. Of course, and again like all social sciences, this is not an easy problem. Over recent decades there has developed a large literature that has begun to seriously look at the issues involved and provide insights into the nature of the problem. This thesis is an attempt to add to this body of work by studying several aspects of voting and electoral competition that have been previously overlooked.

There are many participants and institutions that go into shaping the ultimate policy outcomes within any democratic system. A complete model would incorporate all of these features and highlight their workings and, more importantly, their interdependencies. Unfortunately, such a complete model is currently out of reach. As a consequence, progress has been attempted in two directions. The first approach breaks the problem down into manageable components. The standard demarcation of the problem is into the following three components: (1) voter behavior, (2) candidates and electoral competition, and (3) post-election bargaining. The second approach proceeds, despite the complexity, by combining several of the components into a unified theory. The most successful of these attempts, and to the best of my knowledge the only one that has combined all three components, is the model of a proportional representation election and legislature by Austen-Smith and Banks (1988). However, the price of this success was an oversimplification of the strategic environment at each stage of the process. It is not yet clear which approach, the partial equilibrium or the unified, will lead to the greatest advances. Regardless of this ranking, however, both approaches are valuable and have provided insights into politics and the political process that were otherwise hidden.

In the partial equilibrium approach, each of the three components attempts to understand the behavior of individuals or groups at crucial stages of the electoral process. Models of voting behavior, not surprisingly, focus on voters and the incentives they face. In contrast, the candidates and the policies they represent (and, therefore, implicitly a model of post-election bargaining) are fixed exogenously. These models position the voters as strategic agents and seek to understand their behavior relative to the choice set available to them and the behavior of their fellow voters. The underlying question is the effectiveness of the voting mechanism in making group decisions.

In a similar vein but from a different perspective, models of candidates and electoral competition install the candidates as the strategic agents. These models consider how the competitive pressures of an election impact the policy choices of candidates. In this environment the strategic behavior of voters is not considered and there is no specification for how policy is shaped and implemented after the election.

The third component of the problem is post-election bargaining. In these models the pre-election behavior of both voters and candidates is ignored. Instead the set of victorious candidates and their policies is taken as given. The focus is then on how these candidates, and other relevant groups, interact to determine the policy that is actually implemented.

This thesis does not attempt the grand solution. More modestly, I work in the partial equilibrium world and focus on the first two components of the political problem. I attempt to add to the understanding of voter behavior and candidate competition by studying aspects of these environments that have until this point eluded rigorous analysis. By analyzing strategic behavior in these two environments, I attempt to explain several empirical regularities that are not consistent with results from standard models, and increase our understanding of different voting rules and procedures. I hope that the reader does not consider this limited ambition a shortcoming of this thesis, but rather sees it more appropriately as an indication of the difficulty of the underlying problem. Indeed, perhaps the one thing that political research has taught us is that voting behavior and the political process are as complex, and interesting, as any other social interaction.

This thesis is divided into two parts, reflecting the two research streams to which it contributes. Part I contains two models of candidate competition in which I study the policies that result under several different voting rules and in several different competitive environments. I attempt to explain the number of candidates that are observed in real elections and the policies they choose. In contrast, the focus of Part II is on voters and the decisions they make. I investigate the issue of vote timing and how behavior and outcomes vary with changes in the timing of votes, and characterize the efficiency of different vote timing schemes when used to make group decisions. These research areas, and the place of my work within them, will be described in more detail at the appropriate times. Unfortunately, and quite clearly, these contributions do not complete the grand problem that I initially outlined. However, the fortune in these obvious shortcomings is that the presentation of the results need not be delayed by an outline of future research directions. The main possibilities should be clearly apparent.

Part I

Electoral Competition

Introduction

Models of candidate competition are an attempt to formalize the strategies adopted by electoral candidates in their pursuit of policy outcomes and the perks of office. Of interest is not only which candidates choose to contest the election and the policies they adopt, but, even more importantly, how these platforms translate to policy outcomes. The standard way of understanding this problem is what has become known as the 'spatial model of electoral competition.' This approach was first suggested by Hotelling (1929) and popularized by Downs (1957).

In this 'spatial model' policy choices can be represented by a point in a space. Voters have preferences over these spaces and cast their votes accordingly. Then, with a particular voting rule in effect, these models provide a framework to determine how the different candidates react to this voting pattern and the conflicting ambitions of their competitors. The spatial model has provided enormous insight into the incentives facing candidates for office and the policies they choose. However, there have remained some failings in this development. Firstly, despite the focus of research on elections employing the plurality rule, several well documented empirical regularities have remained unexplained. Rectifying several of these omissions is the purpose of Chapter 1. Secondly, precisely because of the focus of research on elections employing the plurality rule, our understanding of the incentives facing candidates under other voting rules remains woefully incomplete. Chapter 2 attempts to fill up one such hole by studying a model of electoral competition under the run-off rule.

Empirical studies of plurality rule elections have repeatedly observed the following two phenomena: firstly, that only two parties compete in elections, and secondly, that these two parties choose non-centrist policy platforms. The first regularity was even afforded the status of a law by Duverger (1954), and is now commonly referred to as Duverger's Law. Despite these observations, previous theoretical models of electoral competition (with office motivated parties) have been unable to simultaneously explain these two regularities. These models either lead to equilibria involving the entry of more than two parties, or equilibria in which the two parties converge to the same policy platform. I attempt in Chapter 1 to provide an explanation that is consistent with both phenomena. I show that when parties are required to compete simultaneously for many districts, as is the case in most large scale federal elections such as in the United States, the threat of entry induces the first two parties to choose non-centrist policy platforms in order to deter subsequent entry. I characterize the conditions when such a strategy will prove successful, and thus produce two-party non-centrist electoral contests as observed in real plurality rule elections. These conditions can be interpreted as describing the Domain of Duverger's Law.

In Chapter 2, I study a similar model with instead the run-off rule being used to determine the winner. This rule, like many others, has received little attention in the formal literature. The results of this chapter provide some insight into the incentives and policy choices of candidates in such an environment. I find that the subtle change in counting method as compared to the plurality rule yields a significant change to the competitive environment and the resulting policy outcomes. In contrast to the plurality rule results of Chapter 1, I find that a continuum of two-party non-centrist equilibria exist in the traditional single district environment, and that they are robust to simultaneous competition for multiple districts.

Perhaps the most startling finding that can be taken from this work is the significant change in the candidates' strategies and policy outcomes that result when only a subtle change is made to the voting rule. If anything, this confirms the delicate nature of electoral mechanisms, and the importance of their careful implementation. Unfortunately this potential has been ignored all too often in real world situations. Repeatedly, well-intentioned institutional designers have implemented what they thought were innocuous changes only to have them cause far reaching and unintended consequences on political institutions and policy outcomes.[1] Formal work like the following two chapters cannot stop these designers from making costly mistakes. What it can do, at least, is ensure that they have a better idea of what they are

[1]New Zealand and Israel are both excellent examples from the past decade. During the 1990s, New Zealand switched from a single-member district system to one of proportional representation with the aim of broadening representation. However, they were seemingly unprepared for the stifling effect on government and policy formation that such a voting rule would cause. The case of Israel is discussed in footnote 31 of Chapter 1.

getting themselves in for.

Chapter 1 Electoral Competition in Heterogeneous Districts

Abstract

By extending the established theoretical models of electoral competition with entry (e.g., Palfrey (1984)) to incorporate simultaneous competition for multiple districts, I produce a unique two-party equilibrium under plurality rule in which further entry is deterred. Further, this equilibrium requires non-centrist party platforms. These characteristics are consistent with empirical observation, in contrast to those of single district models. Necessary and sufficient conditions for the existence of this equilibrium are then characterized. Taken together, these conditions provide a domain for Duverger's Law. The chapter also sheds some light on how the different levels of elections in the U.S. and other systems relate to each other.

1.1 Introduction

It is a commonly noted feature of plurality rule elections that fewer parties enter and compete than under other electoral rules.[1] This implies that the positions chosen by competing parties in combination with the unique characteristics of the plurality rule are deterring additional parties from entering the election. A second ubiquitous observation from plurality rule elections is that parties choose divergent policy positions.[2] In this chapter, I present a model that simultaneously explains these dual phenomena and shows they are, in fact, complementary. I extend the traditional spatial model of elections to incorporate simultaneous competition for multiple districts and find that there exists a unique equilibrium in which two parties select non-centrist platforms and are able to deter subsequent entry.

As plurality rule is often used to elect members of a legislature, as well as a head of state this extension would seem natural and, in many such circumstances, essential. For example, in the U.S. the main parties compete not only for a single constituency but for 435 Congressional districts simultaneously. Thus, a model of simultaneous competition for multiple districts captures more accurately the problem of platform selection by mass parties.

The model employed here is an extension of that introduced by Palfrey (1984). The candidates are purely office motivated (i.e., they are Downsian) and voting is sincere. To capture the notion of entry deterrence I assume, in contrast to Palfrey, that each potential entrant will enter the election only if it has a positive probability of electoral success. Critically, with this assumption the non-centrist equilibrium of Palfrey no longer exists. Thus, though the equilibrium found by Palfrey explains policy divergence it cannot capture the notion of entry deterrence. In fact, once this assumption is made no pure strategy equilibrium exists in the single district case. This motivates the extension to multiple districts as with some heterogeneity across districts a unique non-centrist two-party equilibrium exists. This is true as long as the degree of heterogeneity doesn't exceed a quantifiable upper bound. These bounds

[1] See, amongst many others, Lijphart (1994).

[2] For the case of the U.S., see Alesina and Rosenthal (1995, chapter 2).

are used to characterize the domain for the existence of a two-party entry deterring equilibrium.

To maintain consistency with the model of Palfrey (1984), I assume that two incumbent parties move simultaneously, and are then followed by potential entrants. Significantly, however, the results I find are not particular to this sequencing and are robust to alternative timing schemes, such as purely sequential or endogenous timing. This is in contrast to single district models in which outcomes depend critically on the timing of moves.

The two-party equilibrium found here is consistent with Duverger's Law. This empirical law claims that not only does plurality rule lead to the entry of fewer parties but, in fact, typically leads to a two-party system. Loosely interpreted the law is well supported.[3] However, the difficulty is that even though plurality rule elections are usually dominated by two parties there is almost always more than two candidates on each ballot.[4] The fact that these additional candidates usually receive a minor share of the votes and do not seem to have a chance of victory is taken as evidence that they are pursuing an alternative agenda and shouldn't be considered as legitimate candidates. This interpretation implies that more than simply a name on the ballot is required before a candidate can be considered legitimate.

In the equilibrium found here, two incumbent parties are able to select platforms such that the entry of further victory seeking candidates is deterred. Thus, even though other non-victory seeking candidates may enter the contest, this equilibrium formalizes the intuition of Duverger's Law. Given this interpretation, my finding of a restricted domain in which the two-party equilibrium exists is appropriate as though there are many positive examples of the law, it does not hold everywhere.

Related Literature

The extension to multiple districts suggested here is only one possible explanation for two-party dominance and non-centrist platforms. There are many variants on

[3]For support of Duverger's Law see the references in Riker's (1982) survey. In Riker's view, "There are indeed counterexamples [to the law], but not, I believe, definitive ones..." (Riker 1982, p.760).

[4]See Lijphart (1994).

the spatial model which employ different assumptions about the capabilities and motivations of both candidates and voters. Due to the inherent difficulty of extracting information on these variables, empirical investigation has been unable to provide definitive suggestions as to which combination of assumptions is the most appropriate. It is likely that reality is some blend of the extremes commonly studied in the formal literature. The hope is that by studying the different extremes of behavior all of the different forces working to produce two-party dominance and non-centrist platform choice under plurality rule can be captured.

Most attempts at explaining the ability of two parties to dominate plurality rule elections have focused on the behavior of voters rather than the candidates. An early example of this approach is Palfrey (1989). The equilibrium he finds shows that if voters are acting strategically then they will focus in on two parties in order to maximize the effectiveness of their votes. This seems to be a plausible explanation. However, empirical investigations of voting behavior conclude that whilst some strategic voting is observed its occurrence is not overwhelming.[5] The model presented here secures more firmly the theoretical understanding of the phenomenon by showing that even if the alternative extreme of voter behavior, sincere voting, is assumed then two-party dominance can still be explained.

Attempts at explaining policy divergence have looked at both the number of parties and their preferences. In the models of Palfrey (1984) and Cox (1987) more than two parties compete and choose non-centrist positions in order to maximize their vote shares. An alternative approach is to assume that candidates are not purely Downsian and instead care about the policy that is enacted. Early papers, such as Wittman (1983) and Calvert (1985), use this assumption in conjunction with uncertainty about the true distribution of voters' ideal points to achieve non-centrist platform choice. More recently, the citizen-candidate models of Osborne and Slivinski (1996) and Besley and Coate (1997) achieve policy divergence and entry deterrence in a full information environment. These papers assume that candidates can only offer their own ideal point as a platform, the opposite extreme to the purely Downsian

[5] See Alvarez and Nagler (1999).

candidates modelled here.

The remainder of the chapter is organized as follows. Section 2 specifies the model and shows why previous models with Downsian candidates and sincere voting have been unable to simultaneously explain two-party dominance and non-centrist platform choice. Section 3 presents the results. Section 4 contains a discussion of the results and the robustness of the model. Section 5 concludes.

1.2 The Model

Most plurality rule elections involve more than one set of candidates and one grand district. Typically competition is for many districts simultaneously. To this end, I construct a spatial model of electoral competition for many districts with entry. The order of play is as follows. There are two incumbent parties who choose their platforms simultaneously and compete in every district. In each district a potential entrant then makes an entry decision, and if he chooses to enter he selects a platform position. The potential entrant will enter in a district only if he has a positive probability of winning that district. The entered parties then engage in the election. This is similar to the model of Palfrey (1984), except there are many districts and each entrant may stay out of the contest.

This extension is incorporated into the spatial model by the following assumption. The issue space is the real line, $\Re$.

Assumption 1.1 *There exists a continuum of districts. In district i the median voter's ideal point is Z_i. The ideal points of district median voters are distributed symmetrically about 0 on the support $[\underline{Z}, \overline{Z}]$, where $\underline{Z} = -\overline{Z}$, according to the cdf G, where $G(\underline{Z}) = 0$ and $G(\overline{Z}) = 1$, and the pdf, g. g is continuous and is either strictly quasi-concave or quasi-convex.*[6] *The distribution of voters' ideal points in district i is given by the cdf $F(x - Z_i)$ for all $x \in \Re$.*

[6]Note that this permits uniform distributions as they are quasi-convex. The restriction to strict quasi-concavity is to rule out particular flat spots in the distribution that may produce multiple weak Nash equilibria.

The assumption of a continuum of districts is, of course, not realistic. However, it has been employed as it captures the effect and intuition of the multiple district scenario whilst avoiding the complexity of calculation associated with a lumpy distribution of district median voters. It is in the same spirit as the assumption of a continuum of voters in the single district case. The intuition of the entry deterring equilibrium found here is not dependent on there existing a continuum, or even a large number of districts. In fact only two distinct districts are needed for such an equilibrium to exist.

In each district there is a continuum of voters with symmetric, single peaked preferences over the issue space. The non-degenerate cdf F and the corresponding pdf, f, have the following properties.

Assumption 1.2 *For all $\alpha < 0$ for which $F(\alpha) > 0$, the function F is strictly increasing on $(\alpha, -\alpha)$.*

Assumption 1.3 *F is continuous and twice differentiable at all points $x \in \Re$ such that $F(x) \in (0,1)$.*[7]

Assumption 1.4 $F(x) = 1 - F(x)\ \forall x \in \Re$.

Assumption 1.5 $f'(x) \geq 0\ \forall x \leq 0$, *and* $f'(x) \leq 0\ \forall x \geq 0$.

These assumptions specify that the distribution of voters' ideal points in each district is symmetric about the median, and that the mass at any point is at least as great as at any point further from the median. This requires f to be quasi-concave. It can be seen that the uniform distribution satisfies these conditions. Assumption 1.2 ensures that there are no gaps in the distribution but without assuming that voter ideal points span all of $\Re$ (that is, voter ideal points can be contained in a bounded interval, for example [-1,1]).

Voters are assumed to be sincere and so each votes for the candidate closest to their ideal point. I will denote the two incumbents as I_1 and I_2, the entrant in

[7]This specification allows for discontinuities in F at only two points: the boundary points of the support of f. This possibility permits the uniform, among others, as a possible distribution.

district i as E_i, and the set of all entrants by E. If a voter is indifferent over any set of candidates then he randomizes over this set.

To differentiate the multiple district approach from the repeated application of the single district result the following assumption is made.

Assumption 1.6 *I_1 and I_2 must each choose a single platform on which they will compete in every district.*

The restriction is somewhat excessive as party nominees typically maintain some degree of freedom in their campaigns. However, it is true that these candidates are associated with the party platform by voters and their nominations depend on party loyalists. Thus, it seems plausible that their flexibility in conveying a platform to the voters in their district is restricted. I take this restriction to the extreme in order to capture this notion fully and simplify the model. However, the results do not depend critically on the extremeness of this assumption. If it is instead assumed that candidates have some freedom around the party platform then the qualitative features of the results should not change. All that is required is that this freedom is not excessive.[8]

Parties are free to locate at any point in the policy space, $\Re$. The incumbent parties have lexicographic preferences with share of districts won on the primary dimension and total vote share on a second dimension.[9] If a party has a set of points that maximizes its share of districts won it chooses the point in this set that maximizes its vote share. If there is more than one point in this set that maximizes a party's vote share then the party randomizes equally over these points. Denote the outcome pair for candidate I_1 by O_{I_1}. Define $M(j)$ and V_j, $j = I_1, I_2, E_i \ \forall i$, to be respectively

[8] More precisely, the two-party equilibrium will exist as long as the freedom of individual candidates to differentiate themselves from the party platform does not exceed $\frac{2}{3}\overline{Z}$. This is in contrast to the multiple district results under the run-off rule presented in Chapter 2. For the two-party equilibrium to persist under the run-off rule the freedom of individual candidates must instead satisfy a lower bound, and in fact be greater than or equal to $\overline{Z}$.

[9] The second dimension is required only to rule out potential equilibria in which neither of the incumbent parties win any of the districts and are unable to move their platforms anywhere such that they do. Without the second dimension such location pairs would constitute equilibria even though it may be asked why the incumbents themselves would enter given they have a zero probability of winning any districts.

the share of districts won by, and total vote share of, party j. In an abuse of notation denote the parties electoral platforms by I_1, I_2, and E, where $E = (..., E_i, ...)$. The outcome function can be written as follows.

$$O_{I_1}(I_1, I_2, E) = (M(I_1|I_1, I_2, E), V_{I_1}(I_1, I_2, E))$$

The outcome for I_2 is defined analogously. Strict (weak) preference for one outcome over another is denoted in the usual way by the binary relation $\succ$ ($\succeq$), where $A \succ B$ represents the situation in which outcome A is strictly preferred to outcome B.

Each entrant is assumed to have the single district analogue of these preferences; the first dimension is probability of victory and the second dimension is vote share. Given the positions of the incumbents, there may not exist an optimal location choice for each entrant. This technicality arises when an entrant attempts to maximize his vote share over the set of points that maximize his probability of winning the district. The probability of winning for any E_i can only take on a finite set of values (as there are only three candidates and voting is deterministic) and so a set of maximizers over this dimension can always be found.

A variant of the limit equilibrium concept introduced by Palfrey (1984) is employed to deal with this problem. If a vote maximizing point doesn't exist then the entrant 'almost' maximizes his vote share when choosing from the set of points which maximize his probability of winning. A perturbed game is defined for each ε, where ε is how close each E_i comes to maximizing his vote share. An equilibrium is defined as any pair of strategies for I_1 and I_2 which are best responses to each other for an infinite sequence of the perturbed games, with the perturbation approaching zero in the limit. The difference between this equilibrium and that of Palfrey is that his candidates simply maximize vote share and their preferences are representable by a utility function. Here the candidates have lexicographic preferences and such a general representation is not possible. Consequently, the definition of equilibrium must remain in terms of primitive preferences. Technically the definition introduced here is stricter than Palfrey's, though in the single district case (where both definitions could be used) the results are unaffected by this additional requirement.[10]

Denoting the winner of a district i by W_i, the set of points that maximize E_i's probability of victory is defined as follows.

$$X_i(I_1, I_2) = \arg\max_{x\in\Re}\{prob(W_i = E_i|E_i = x)|I_1, I_2\}$$

If $\max_{x\in\Re}\{prob(W_i = E_i|E_i = x)|I_1, I_2\} = 0$ then E_i does not enter the election and $E_i = \emptyset$. If not, then the set of points that E_i equally randomizes over, for a given ε, is given by $C^{\varepsilon}_{E_i}$, where,

$$C^{\varepsilon}_{E_i}(I_1, I_2) = \{E_i \in X_i(I_1, I_2)|V_{E_i}(I_1, I_2, E) > V_{E_i}(I_1, I_2, y) - \varepsilon, \forall y \in X_i(I_1, I_2)\}$$

Let $C^{\varepsilon}_{E}(I_1, I_2) = \left(..., C^{\varepsilon}_{E_i}(I_1, I_2), ...\right)$. Anticipating these entry decisions the expected outcome for the incumbents, given their own locations, is the expectation over $C^{\varepsilon}_{E}(I_1, I_2)$.

Definition 1.1 *A pair of locations, $\{I_1, I_2\}$, is a strict limit equilibrium if,*

(a) for every $y \neq I_1$, there is an $\varepsilon(y)$, such that for all $E' \in C^{\varepsilon(y)}_{E}(y, I_2)$ and $\tilde{E} \in C^{\varepsilon(y)}_{E}(I_1, I_2)$, $O_{I_1}(I_1, I_2, \tilde{E}) \succ O_{I_1}(y, I_2, E')$. And,

(b) for every $w \neq I_2$, there is an $\varepsilon(w)$, such that for all $E' \in C^{\varepsilon(w)}_{E}(I_1, w)$ and $\tilde{E} \in C^{\varepsilon(w)}_{E}(I_1, I_2)$, $O_{I_2}(I_1, I_2, \tilde{E}) \succ O_{I_2}(I_2, w, E')$.

It is possible that if $X_i(I_1, I_2)$ is not a singleton then an entrant will randomize over one or several intervals. For any policy position, $x \in \Re$ and $\varepsilon(x) > 0$ such that $\varepsilon \to 0 \Rightarrow \varepsilon(x) \to 0$, define the following intervals: $x^{+} = (x, x + \varepsilon(x))$, $x^{-} = (x - \varepsilon(x), x)$, and $x^{-+} = (x - \varepsilon(x), x + \varepsilon(x))$. If $I_1 \leq I_2$ then every entrant, if they choose to compete, will locate in some subset of the intervals I_1^{-}, I_2^{+}, and z^{-+}, where $z \in (I_1, I_2)$.

As the parties have no ideological motivation in the selection of their platforms it is obvious that any equilibrium found will point to another equilibrium in which

[10]See Callander (1999).

the two incumbent parties simply switch positions. Any pair of such equilibria will be considered to be the same, and so constitute just one equilibrium.

$\underline{Z} = \overline{Z}$: The Single District Case

If $\underline{Z} = \overline{Z}$ then all districts are homogeneous and the analysis collapses to the traditional single district case.[11] In a separate paper, Callander (1999), I prove that in this case no pure strategy equilibrium exists.[12,13] The intuition behind this result is that given any non-centrist pair of platforms for the incumbents a small enough deviation can always be found such that the entrant chooses to stay out of the election and the deviating incumbent wins. This possibility highlights a critical feature of plurality rule elections: that entry on the flank cannot punish both incumbents simultaneously. If the entrant locates on the flank he secures all of his votes at the expense of the incumbent he outflanked. Critically, the votes won by the other incumbent are unaffected. Thus even though the entrant may win a substantial share of the votes (when the incumbents are close to the median) he can't beat both incumbents and can't win the election.

Palfrey (1984) achieves a non-centrist equilibrium in this framework but requires that the third candidate always enters.[14] Thus, his equilibrium cannot capture the notion of entry deterrence. The nonexistence of an equilibrium in the single district case motivates the extension to multiple heterogeneous districts as once this extension is made the existence problem will be overcome and a unique, non-centrist, entry deterring equilibrium will exist.

The Multiple Districts Literature

There have been few papers which have considered the issue of multiple districts. The first investigation of multiple districts was by Hinich and Ordeshook (1974) in a

[11] The primary dimension of the objective function then collapses to the probability of victory.

[12] Osborne (1993) considers this model for when the candidates move simultaneously or whenever they wish.

[13] Feddersen, Sened, and Wright (1990) extract an equilibrium from this framework by assuming voters are able to coordinate their decisions. However, the equilibrium they find involves centrist platforms and thus can't explain policy differentiation.

[14] I also prove in Callander (1999) that this equilibrium still exists and is unique for the more general policy space, party objective function, and equilibrium notion employed here.

study of the electoral college. Hinich and Ordeshook were interested in distortionary effects of the electoral college in comparison to a direct vote for the President. They proved the extension of the single district case, that with two candidate competition both candidates would converge to the median of the median district. This result makes two candidate competition in the multiple district case look almost identical to that in the single district case (though maybe with a different convergent point). Further work has been done by Austen-Smith in a series of papers. Austen-Smith (1981) is the most similar to the model presented here as it is assumed that parties must choose a unique platform which is applicable for candidates in all districts (Assumption 1.6 here). The question of entry and entry deterrence is not considered and parties compete on both distributive and policy dimensions. Consequently, the results and intuition obtained are significantly different from those presented here.

In a recent paper, Osborne (2000) constructs a model similar to this but with different motivation. In his model, parties compete for a single district but are uncertain as to the true distribution of ideal points. This uncertainty is symmetric across parties and consequently a potential entrant can't enter selectively. Thus the equilibria obtained are significantly different and, in fact, the domain of the two-party equilibrium that he finds is disjoint from that found here.

1.3 Results

If competition is for only one district then there are many pairs of locations for the incumbents such that entry is deterred. The incumbents may be located symmetrically or even asymmetrically around the median voter and deter entry, as long as they are not too far from the median nor too asymmetric. In fact, the nonexistence of an equilibrium in the single district model is a consequence of the abundance of such location pairs. This is because the incumbents, regardless of where they are located (as long as they are deterring entry), can always find a deviation such that entry is still deterred and they are better off (see Callander (1999) for the proof of this result).

However, when the incumbents are required to compete simultaneously for multiple heterogeneous districts many of the location pairs which deter entry in one district permit entry in other districts. This follows from the fact that if the incumbents are located symmetrically in one district then they must be located asymmetrically in dissimilar districts (as they can only choose one policy platform for all districts). As the heterogeneity of districts increases then the asymmetry of the incumbents' locations in some districts must also increase. The following lemma shows how extreme the heterogeneity can be for any given positions of the incumbents (that aren't too far apart) such that entry is deterred. Effectively, I find that the incumbents must be spaced twice as far apart as are the most extreme district median voters if they wish to prevent entry in every district. Define $Z' = (2Z^*, -2Z^*)$, where $F(Z^*) = \frac{1}{3}$. All proofs are gathered in the appendix.

Lemma 1.1 *For $I_1, I_2 \in Z'$ and $I_1 < I_2$ the districts won by both incumbents combined are those with median voters in the interval $D(I_1, I_2) = [\frac{3I_1+I_2}{4}, \frac{I_1+3I_2}{4}]$.*

The logic of this result can be seen in Figure 1.1 which depicts a three district example. The districts are evenly spaced with median voters at $\underline{Z}$, 0, and $\overline{Z}$. For the district with median at $\overline{Z}$ the density at 0 is the same as at $2\overline{Z}$. Thus, if $I_2 = 2\overline{Z}$ and $I_1 = 2\underline{Z}$ then an entrant locating in I_2^+ (for small enough ε) is just defeated by I_1 whose vote share for this district is given by $F\left(0 - \overline{Z}\right)$. For a district with median voter more extreme than $\overline{Z}$ this is no longer true. An entrant in I_2^+ squeezes I_2 in the center and just defeats I_1. Therefore, this critical point (and the one at $\underline{Z}$) defines the bounds of the lemma.

This example expands naturally to a continuum of districts. It is easy to see from the figure that if the incumbents are deterring entry in the districts with median voters at $\underline{Z}$ and $\overline{Z}$ then they would also be deterring entry in any district with the median voter between these bounds. Thus, it is not the number of districts but rather the heterogeneity of the district medians that determines whether entry can be deterred in all districts simultaneously. This lemma will prove critical to the following results.

For the statement of the propositions I will need the following condition.

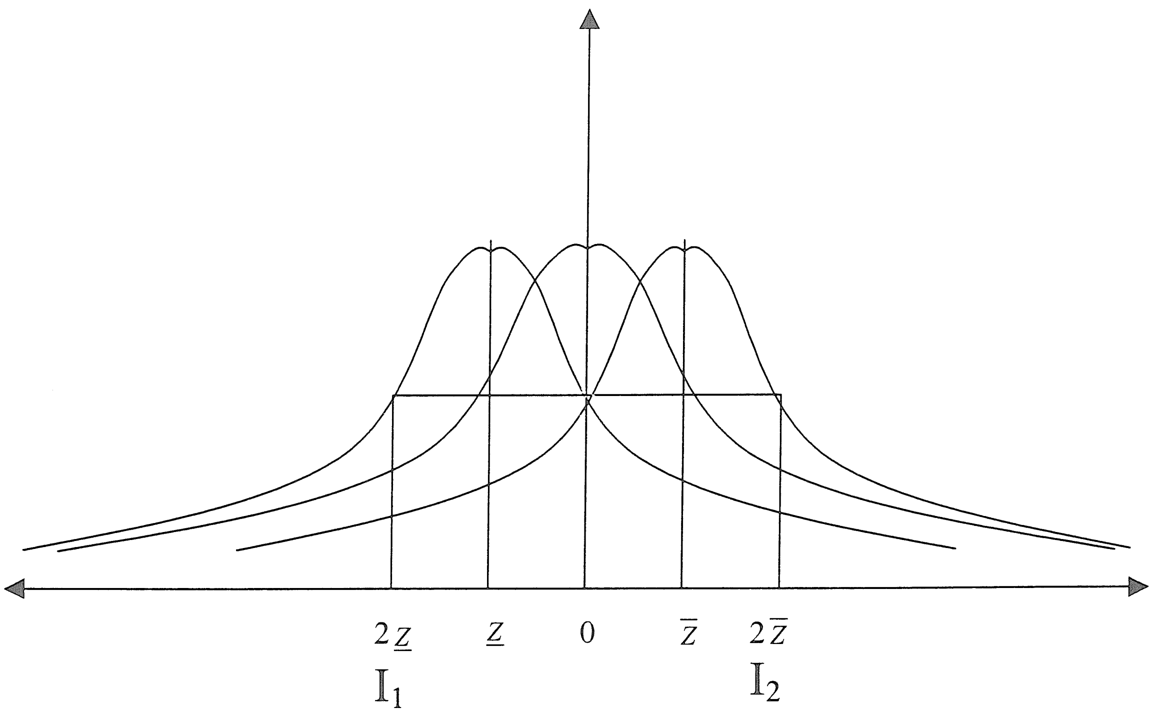


Figure 1.1: Competition for Heterogeneous Districts

Condition 1 $g(\underline{Z}) \geq \frac{2}{3}g(0)$.[15]

This condition ensures that the weight of districts with median voters at the boundary of the distribution is enough such that the incumbents do not have the incentive to abandon them to entrants by deviating inwards in order to win more districts at the center of the distribution.[16] Recalling that g is symmetric $\left(\underline{Z} = -\overline{Z}\right)$, then it must also be true that $g\left(\overline{Z}\right) \geq \frac{2}{3}g(0)$.

The following proposition reflects the main insight of the chapter. It shows that when parties are required to simultaneously compete for multiple heterogeneous districts there exists a unique equilibrium which is consistent with the dual phenomena of entry deterrence and non-centrist platform choice. Thus, it is only in the case of competition for a single district that such an equilibrium doesn't exist.

Proposition 1.1 *Suppose* $0 > \underline{Z} > Z^*$, *where* Z^* *satisfies* $F(Z^*) = \frac{1}{3}$. *Then if Condition 1 is satisfied the unique equilibrium is given by,* $\{I_1, I_2\} = \{2\underline{Z}, -2\underline{Z}\}$, $E_i = \emptyset\ \forall i$ *(do not enter).* $M(I_1) = M(I_2) = \frac{1}{2}$, $M(E) = 0$.

In the single district case the incumbents have an incentive to deviate towards the center and win the election. This incentive still exists in the multiple district case. However, the incumbents reach a point where further deviation to win central districts will allow entry in the districts with the most extreme median voter, as in those districts the incumbents are located too asymmetrically (as seen from Lemma 1.1). Condition 1 ensures that the share of districts lost on the edge by deviating inwards outweighs the share won in the center.

Significantly, these results are not dependent on the particular timing scheme assumed here. This is in contrast to the single district case in which timing changes have a significant effect on the set of equilibria to the candidate competition game.[17]

[15] Note that this condition only restricts strictly quasi-concave distributions and places no restrictions on quasi-convex distributions of districts.

[16] The $\frac{2}{3}$ arises because if an incumbent deviates towards the median then he wins an interval of districts in the center which is $\frac{2}{3}$ the length of the interval of districts he loses on the flank. Thus, to ensure that this deviation is not profitable the density at the center can be no greater than $\frac{3}{2}$ the density on the flank.

[17] This effect can be seen by comparing the results of Osborne (1993), who compared simultaneous

The robustness of the results presented here follows from two features of multiple district electoral competition: (1) a single party is unable to locate at the median simultaneously in heterogeneous districts (which explains why the single district case is affected differently), and (2) if the first party to enter is located close enough, but not at the median in a district then a second party can win these districts and prevent further entry. This implies that the first party to enter will locate with the expectation of additional entry. Simple arguments relying on Lemma 1.1 can then be used to show that the above equilibrium, as well as those in subsequent results, must still exist if the timing scheme were instead endogenous or purely sequential.[18] Thus, the results presented here are not dependent upon the particular timing structure employed.

In equilibrium, the incumbents win half of the districts each. They tie in the central district, which is then decided by randomizing. However, as there is a continuum of districts, this central district has a weight of zero and doesn't affect the proportion of districts won by each of the incumbents. If the continuum of districts is thought of as the limit of a finite distribution of districts, then it is only in the limit that the winner of the central district does not achieve a majority and win government outright.[19]

Only occasionally are legislatures observed where the seats are evenly divided between the two major parties, or where they are separated by only one seat. It is quite normal to observe legislatures where one party holds a significant majority. Consequently, it would be desirable if a theoretical model could produce such uneven seat allocations as an equilibrium. Obviously, the equal proportion of seats for the incumbents predicted by this model is a direct consequence of the symmetry of the structure and this would need to be dropped to produce an asymmetric outcome.

timing with endogenous timing, with the single district results described in this chapter (that also extend to a purely sequential structure).

[18]Assume that $I_1 \leq 0$. Lemma 1.1 implies that I_2 will be chosen such that $\frac{I_1+3I_2}{4} = \overline{Z}$. Anticipating this choice I_1's optimal strategy is to locate at $2\underline{Z}$, and thus the equilibrium will be that found in Proposition 1.1.

[19]Of course, that the result of Proposition 1.1 still constitutes an equilibrium with only a finite number of districts remains to be proven.

This is possible in the multiple district framework presented here as the proof of Proposition 1.1 does not necessarily rely on the symmetry of g. Indeed, as long as an analogue of Condition 1 holds (Condition 1A below) then the equilibrium depends solely upon the width of the distribution of districts and not on the shape of the distribution (e.g., the mean or the median). This is an interesting result as it is not automatic that asymmetric distributions produce asymmetric outcomes for the parties.

For example, if asymmetric distributions are incorporated into models that predict party convergence then the parties still converge to the median (of the median district in the multiple district case), though this may no longer be in the geographic center of the distribution. Thus, even though there may be a different set of platform choices by the parties they still receive symmetric outcomes. In Palfrey's model it is possible to produce asymmetric outcomes for the incumbents for rather special distributions. However, this implies that one incumbent has no chance of victory and therefore there is no clear reason why this incumbent would enter the election. Consequently an interpretation of asymmetric outcomes in the single district framework is difficult to develop. However, in the multiple district framework such asymmetric outcomes are easily conceptualized. Even though the minor incumbent party has no chance of winning a majority, it still wins some districts contested and thus secures a voice in the legislature. Entry in this instance could also be justified by the fact that some members of the losing party still gain personally by winning their own district.

To characterize the equilibria for an asymmetric distribution of districts, Condition 1 will need to be generalized and strengthened. To maintain tractability, I shall continue to assume that $\underline{Z} = -\overline{Z}$, though this too is not required.

Condition 1A $g(x) > \frac{2}{3}g(y)\ \forall x, y \in [\underline{Z}, \overline{Z}]$.[20]

The tightening of this condition is required to rule out certain flat spots in quasi-convex distributions. Such an additional restriction was not required in the statement

[20]This condition simply states that the density of g doesn't vary excessively. This will ensure that profitable deviations for the incumbents do not exist.

of Condition 1 as symmetry ensured that even if such flat spots existed they would not cause a problem. This tightening is overly strong. Consequently, whereas Condition 1 was sufficient and necessary for the result of Proposition 1.1 to hold, this condition is sufficient for the following result but not necessary.[21] This leads to a generalization of Proposition 1.1.

Proposition 1.1A *Suppose* $0 > \underline{Z} > Z^*$, *where* Z^* *satisfies* $F(Z^*) = \frac{1}{3}$, *and that* g *is not necessarily symmetric. Then if Condition 1A is satisfied the unique equilibrium is given by,* $\{I_1, I_2\} = \{2\underline{Z}, -2\underline{Z}\}, E_i = \emptyset\ \forall i$ *(do not enter).* $M(I_1) = \int_{\underline{Z}}^{0} g(x)dx$, $M(I_2) = \int_{0}^{\overline{Z}} g(x)dx$, $M(E) = 0$

It can be seen immediately that unless $G(0) = \frac{1}{2}$ then $M(I_1) \neq M(I_2)$ and one of the incumbent parties will hold an outright majority. Thus, the equilibrium involves the selection of party platforms such that one party is guaranteed of winning a majority of the districts. The existence of such equilibria could be used to explain elections where one party is predicted to win a clear majority and does so, and where the losing party does not seem to have a platform that could win a majority of the seats.[22] This is consistent with a common analyst observation that a party has 'captured the middle ground.' A distribution of districts that is skewed to one side would produce such an outcome.

An interesting technical detail of this result is that the core of the candidate game (the median of the median district) can vary with the distribution of district median voters, but the equilibrium doesn't change. In fact, the equilibrium only depends on the width of the distribution. This leads to the question of whether the equilibrium would still exist even if the core didn't. As the current model is restricted to one dimension, a core always exists and so an answer to this question cannot be immediately attained. However, when spatial competition is over multiple dimensions a core does not generally exist and consequently neither does an equilibrium. Therefore, the

[21] The necessity of Condition 1 for the equilibrium of Proposition 1.1 is proven in Proposition 1.3.

[22] Potentially we could also explain such an outcome if we considered a dynamic model in which the distribution of districts changed from election to election but parties were restricted in changes to their platforms. The purpose of the result here is to show that such an uneven outcome is also possible in a single election model with parties completely free to select their platforms.

possibility that simultaneous competition for many districts may lead to equilibrium existence in many dimensions would seem to be worthy of further investigation.

In addition to this assumption of symmetry, the statement of Proposition 1.1 required several further restrictions. As these restrictions may not always be valid, it is of interest to determine the equilibria of the game when they are violated. The following two propositions consider these possibilities. Proposition 1.2 assumes the heterogeneity of districts to be greater than permitted in Proposition 1.1, and Proposition 1.3 considers the situation when Condition 1 is not satisfied. Significantly, in both of these cases any equilibrium must involve the entry of more than two parties.

Proposition 1.2 *Suppose $\underline{Z} \leq Z^*$. Then if an equilibrium exists it must involve the entry of more than two parties.*

In this situation, the dispersion of districts is too broad for the incumbents to compete successfully in all of them. To prevent successful entry at a point between themselves in the central district, the incumbents must leave open the possibility for successful entry in the extreme districts by locating too asymmetrically. The result here is stronger than what is stated: in fact, the two incumbents can't prevent entry whether they are, or are not, in equilibrium. As it is the equilibria that are of interest, the result has been stated in its weaker form.

Proposition 1.3 *Suppose $0 > \underline{Z} > Z^*$, where Z^* satisfies $F(Z^*) = \frac{1}{3}$, and assume that Condition 1 is not satisfied.*[23] *Then if an equilibrium exists it is unique and is given by, $\{I_1, I_2\} = \{2Z^\#, -2Z^\#\}$, where $Z^\# < 0$ and satisfies $g(Z^\#) = \frac{2}{3}g(0)$. Entrants enter and win districts with median voter's ideal points in the intervals, $[\underline{Z}, Z^\#)$ and $(-Z^\#, \overline{Z}]$. If g is concave then such an equilibrium always exists.*

With Condition 1 violated, each incumbent has an incentive to deviate inwards to win districts in the center from the other incumbent, even though this involves giving up the extreme districts to entrants. As with the equilibrium in Proposition 1.1, the incumbents still have equal shares of expected district wins, but now neither party

[23]Therefore, g must be strictly quasi-concave (as all quasi-convex distributions satisfy Condition 1).

will hold a majority. The equilibrium is given by the point where further inward deviation involves more districts lost on the edges than gained in the center. The continuity of the pdf g ensures that such a point exists.

This is very similar to the equilibrium when Condition 1 is satisfied. Condition 1 simply ensures that the critical point is reached before any entry occurs. Therefore, Condition 1 can be seen as a necessary condition for an equilibrium to involve only two parties. Though the districts abandoned on the edges by the incumbents could be won by a different party entering in each district (effectively independents), there could be as few as two parties entering and winning arbitrarily close to one half of these districts each. To determine the final party structure in this instance, a more extensive model of entry would need to be formalized. As two-party outcome structures and entry deterrence are the primary concern, this question will not be explored any further here.

As with Proposition 1.1, the assumption that g is symmetric could be dropped here to produce an equilibrium involving entry and asymmetric seat shares for the incumbent parties. Indeed, particular g functions could be found to produce any variety of multiple party equilibria, for example involving entry only on one flank.

This location pair may not constitute an equilibrium for non-concave g functions if there is too much district share that is lost to entrants. That is, if the district shares of the incumbents is so small that they each have an incentive to deviate from the prospective equilibrium to the outside of the other incumbent as they can win a greater share of the districts on the flank. If g is concave then the density on the flanks is small enough such that these deviations are not profitable and there exists an equilibrium.

A necessary condition for the existence of a two-party equilibrium can also be extracted from Proposition 1.2. This result showed that for a two-party equilibrium to exist the heterogeneity of district medians can't be too great. The following condition captures this requirement precisely.

Condition 2 $[\underline{Z}, \overline{Z}] \subset [Z^*, -Z^*]$.

Therefore, taken together Conditions 1 and 2 form a necessary and sufficient condition for an equilibrium to involve at most two parties. As such, these two conditions can be interpreted as the limit of Duverger's Law. If both conditions are satisfied then two parties are able to deter subsequent entry and the law will be satisfied. If either condition is violated then there would be entry and the law would not apply. The following theorem summarizes these results.

Theorem 1.1 *Suppose Assumptions 1.1-1.6 are satisfied. Then together Conditions 1 and 2 constitute a necessary and sufficient condition for a two-party entry deterring equilibrium to exist.*

As such, the requirement for cumulative density functions G to satisfy Conditions 1 and 2 can be seen as characterizing the domain of Duverger's Law. Outside of this domain the law would not be expected to hold. This is an appropriate result because to explain an empirical law such as Duverger's, that doesn't hold universally, a theory that predicts a restricted domain of applicability would be expected, and indeed desired. Hopefully a recourse to empirics will indicate whether this is the correct restriction.

1.4 Discussion

1.4.1 The Model

In many respects the model assumed here is very specific. In this section, I will attempt to defend the validity of several of these restrictions and discuss the robustness of the results to their variation.

A primary restriction is that the parties must move in a fixed order: the incumbents simultaneously choose their platforms, followed by the entrants. This structure was chosen to highlight the notion of entry deterrence. The order of play shows clearly how the platform selection of the incumbent parties deters potential entrants from contesting the election. As real elections are a repeated process, the option for an entrant to move after the incumbents would seem almost necessary. This, of course, does not imply that there are no other orderings that satisfy these desires. Alternative specifications require the timing of play to be completely sequential or, perhaps most appropriately, endogenous.[24] Significantly, the results presented in the previous section are robust to these alternative timing schemes.[25]

A further feature of the model is that there exists an asymmetry between the parties. This asymmetry leads to the natural question: why do the incumbent parties compete in districts in which they do not win, and why don't potential entrants do the same and enter on a large scale? Unfortunately, the empirical literature provides no definitive suggestions as to party motivations or intentions. When competition is for multiple districts, this uncertainty is magnified. The structure employed here would seem to be an acceptable possibility for several reasons. Foremost among these is that it captures the steady state competitive structure that real world observations are actually describing. Duverger's Law describes the empirical regularity of two large parties deterring subsequent entry. For these observations there does not exist any other mass political parties, for if groups do not enter candidates in the election then they are, by definition, not political parties. Instead, the large parties face the threat of entry from opportunistic individuals. The strategic ability of the large parties to deal with this ongoing threat is captured by the equilibrium and timing scheme employed here.

Given the asymmetry of parties, it may be questioned whether the assumption of sincere voting is appropriate. However, such a criticism ignores the fact that sincere voting is a behavioral assumption and not a dictate of rationality. Therefore, it would not seem any more or less inappropriate in its use here than it would in any other environment. In fact, even if we look at this voting decision from a rational choice perspective, then voting for an independent candidate may be utility maximizing. This follows because even though an independent member of the legislature will not be a member of the majority party (if one exists), he holds the possibility for enormous

[24]Another option, simultaneous choice, does not capture entry deterrence (as described above).

[25]See the discussion on page 22 for details.

influence should he hold the balance of power in a minority government. And as he is not part of a party machine he can use this power to directly benefit the members of his constituency. This potential payoff may exceed that from electing a member of the majority and make a vote for an independent worthwhile.

An additional simplification of the model is that there is only one potential entrant in each district. This assumption is made in order to simplify the analysis. If there were many potential entrants for a single district, then the entry decisions of these parties would be interrelated and would require a more complicated stage game to be specified. The intention, which is maintained by the assumption, is that if a single party could enter and win a district then the incumbents lose that district.

This assumption will not require as many entrants as may be thought. Even though only one candidate will enter in any district it may be that one minor party enters in many districts. In most instances, one entrant, with one platform, will be able to win many districts from the incumbents. Indeed, for the only equilibrium result specifying entry, Proposition 1.3, only two entrants are required to secure all but an arbitrarily small number of the districts lost by the incumbents.[26] Consequently, this framework is rather general and is consistent with several types of political entrant. It can be seen as covering the entry of multiple independents into the legislature, or the creation of regional or single issue based parties which pick off certain sections of the electorate (a good example of this type of entry may be found in Canada where regional parties dominate).

Further, the restriction to one entrant in each district is not as restrictive as it may seem. In fact, if it were assumed that potential entrants were strategic and conscious of further entry in districts they attack then as long as the incumbents are on either side of the median, and entry is possible, it can be shown that one entrant can secure victory in a district and prevent the successful entry of another party.[27,28]

---

[26] This is achieved by the entrants locating at points arbitrarily close to each incumbent.

[27] For example, if in the central district $|I_1| > |I_2|$, and successful entry on the right flank is possible, then $E = I_2 + \delta$, where $1 - F(\frac{I_2+E}{2}) > F(\frac{I_1+I_2}{2})$ but $1 - F(E) < F(\frac{I_1+I_2}{2})$, secures victory for the entrant and prevents further entry.

[28] The assumption of only one entrant in each district allows me to deal with problematic situations in which the incumbents are on the same side of the median, and so wouldn't be expected to win

Alternatively it could be assumed that there is only one large entrant who is, like the incumbents, constrained to a single platform (Assumption 1.6). If this entrant only enters in districts it can win then it can easily be seen from Lemma 1.1 that the two-party entry deterring equilibrium of Proposition 1.1 still exists. Given the entrant moves last, the ability to enter selectively would seem appropriate.

It is assumed in the model that parties seek primarily to maximize their share of districts won. There are many alternative specifications of the party objective function that could be employed. Perhaps the most plausible would be to maximize the probability of winning government. Unfortunately, as governments can be formed with a minority of seats, or by forming a coalition of parties, a complex model of government formation would need to be incorporated for this assumption to be used. Another alternative would be to assume three dimensions of lexicographic preferences with the first dimension being the probability of winning a majority of the districts and the other two dimensions remaining the same. It is easy to see that this would not alter the results of the model as the model is one of perfect information. Thus any deviation that improves a candidate's seat share must also weakly improve its probability of winning a majority, and the same deviations prove the results for this different specification of preferences.

1.4.2 The Results

The results of this multiple districts model can also provide some insight into the relationship between U.S. Congressional elections and Presidential elections. To maximize performance in the House elections and to preclude entry of a third party, each of the two incumbent parties must choose a non-centrist platform. However, the Presidential candidate of each party competes in only one district, the grand district (with mean zero in this symmetric framework), and so would like to move towards this center to maximize his vote in the Presidential race. However, his party is constrained to its non-centrist platform. Therefore, to achieve any centripetal movement

the district, but where one entrant can't prevent subsequent entry.

a Presidential candidate must try and detach himself from his party so that he can move towards the center without disrupting the equilibrium for the House elections. One obvious way to achieve this objective would be on non-policy issues (as they are party platform constrained on policy issues). This can be seen to lead to the cult of personality phenomena in Presidential races. Personality traits are one way for a candidate to make himself seem more central without dragging his party with him.

In fact, these incentives for detachment from the party base are applicable to all candidates, including Senators and district candidates, who want to move towards the median in their given district. Presidential and Senatorial candidates can usually achieve this detachment more effectively than the district candidates, primarily because they are more visible.[29]

This explanation for the cult of personality campaigns so evident in U.S. elections can also be used to explain why such campaigns are not as evident in other single member district elections, such as in Britain and Australia.[30] In those countries, the Prime Minister is elected indirectly by voting for his candidate in each district. Thus leaders of the incumbent parties, the Prime Ministerial candidates, each maximizes his probability of success by maximizing the number of districts that his party wins. And this is achieved by sticking firmly to the non-centrist party platform.[31]

This analysis can also be used to consider the phenomenon of third party candidates in Presidential elections. For a wide dispersion of median voter points the prediction is that the two incumbent parties are also widely spaced. If the Presidential candidates cannot achieve detachment from their party platform, or cannot

[29]To be precise, the Presidential candidates would not attempt to move completely to the center as they are actually competing in a multiple district election with each district representing a state. As Hinich and Ordeshook (1974) pointed out, the candidates would attempt to move to the median of the state that contained the median electoral college vote.

[30]For a discussion and review of this topic, with particular reference to these three countries, see Crewe and King (1994).

[31]Israel is an interesting example of how direct versus indirect election of the leader of the Government can have a significant effect on the political landscape. Electoral changes introduced for the 1996 elections added an additional ballot to the Knesset elections in order to directly elect the Prime Minister. Previously the Prime Minister had been elected indirectly as in other parliamentary systems. This apparently innocuous change has had a dramatic impact on Israeli politics. As the Knesset elections employ proportional representation, the results of the model presented here are not directly applicable. For a full account of the effects of this change on Israel, see Arian (1998).

do it very well, then there will exist a large gap between the positions of the two incumbent party Presidential candidates. It is potentially this hole that the third party candidates have tried to exploit. However, the model also predicts that for an entry precluding equilibria the two incumbents are located no further apart than $[2Z^*, -2Z^*]$. As this isn't wide enough for an entrant to steal the central district it isn't wide enough for a third candidate to steal the Presidential election. The third candidate will, however, potentially receive a large share of the votes even though they have no chance of victory. This prediction is also consistent with history as third party candidates have received a surprisingly large vote share but have never been victorious.[32]

This thinking leads to the question of why doesn't one of the incumbent parties enter a stooge near the other incumbent's platform to break up the oppositions vote and so ensure victory for themselves? Staying strictly within the framework presented here it would be hard to answer that they wouldn't. The constraint would be finding a credible independent candidate, and those that exist would be unlikely to stoop to such behavior to aid a party that they, by definition of being an independent, have little affiliation with. Consequently, such behavior has been ruled out as unachievable (not to mention unethical).

Another empirical fact is that district members in the U.S. House express far greater vote independence than do the equivalent members in, for example, Britain's House of Commons.[33] It could be conjectured that this greater independence is reflective of the ability of the individual members to detach themselves to some degree from their party's platform. It is possible that this extra ability allows the U.S. incumbent parties to support a wider dispersion of median voter points whilst still precluding entry.[34]

A further point which the model predicts that is consistent with observation is that the dispersion of median points produces some districts that are safely in the hands

[32]Smallwood (1983, p.13).

[33]Cain, Ferejohn, and Fiorina (1987, p.43).

[34]Obviously if the candidates are flexible in their platform choices, then the equilibrium itself may change. Though, as long as the freedom of individual candidates is not excessive the same substantive results are obtained.

of one party, others safely in the hands of the other party, and some districts that are fought for fiercely. This is a direct consequence of the constraint that parties must choose only one platform to compete with in every district regardless of its particular distribution of voters. This result can be seen as a formalization and explanation of what Robertson (1977) categorized as marginal and safe seats.

1.5 Conclusion

By extending the established theoretical models of electoral competition with entry to incorporate simultaneous competition for multiple districts, I produce a unique two-party equilibrium under plurality rule in which further entry is deterred. Further, this equilibrium requires non-centrist party platforms. These characteristics are consistent with empirical observation, in contrast to those of single district models. Necessary and sufficient conditions for the existence of this equilibrium are then characterized. Taken together these conditions provide a domain for Duverger's Law. The chapter also sheds some light on how the different levels of elections in the U.S. and other systems relate to each other.

1.6 Appendix

As is standard, all proofs will proceed by showing there exists profitable deviations from any candidate locations other than those claimed to constitute equilibria. I will denote the locations under consideration by I_1, I_2, and E, and any deviation with a tilde (e.g., $\tilde{I}_1$). For simplicity, if the arguments of a function are I_1, I_2, and E then they will be omitted. In an abuse of notation let $E_i = I_j^+$ denote an entrant in district i locating arbitrarily close to the right of incumbent party j. Further, represent an incumbent j's vote share in district i by $V_{I_j}(i)$. WOLOG assume that if $I_1 \neq I_2$ then $I_1 < I_2$, and define $Z' = (2Z^*, -2Z^*)$.

1.6.1 Proof of Lemma 1.1

Note that all of the districts in this interval may not exist for a given G (i.e., existence requires $g(.) > 0$). The lemma is proven by showing that successful entry is not possible in these districts, and only these districts.

Case 1: Entry between the incumbents.

For any district l, with median voter z_l, the vote share of an entrant is given by $V_{E_l} = F(\frac{I_2+E_l}{2} - z_l) - F(\frac{I_1+E_l}{2} - z_l)$ and the length of this interval is $|V_{E_l}| = (\frac{I_2-I_1}{2})$. As $I_1, I_2 \in (2Z^*, -2Z^*)$, then $(I_2 - I_1) < -4Z^* \implies (\frac{I_2-I_1}{2}) < -2Z^*$. By the definition of Z^* it must be that $V_{E_l} < \frac{1}{3}$ and, therefore, successful entry is not possible.

Case 2: Entry on a flank.

Let district l have median $z_l = \frac{3I_1+I_2}{4} - \delta_l$, where $\delta_l \in [\frac{I_1-I_2}{4}, \infty)$. If $E_l > I_2$ then $V_{I_1}(l) > V_{E_l}$ as $|I_1 - z_l| < |E - z_l|$. Therefore, for entry to be successful in any district l it must be on the left flank.

Suppose $E = I_1^-$. V_{E_l} is bounded by $F(I_1 - z_l)$ as $\varepsilon \to 0$. This implies that, as $\varepsilon \to 0$, $V_{E_l} \to F(I_1 - \frac{3I_1+I_2}{4} + \delta_l) = F(\frac{I_1-I_2}{4} + \delta_l)$ from below. As F is symmetric $F(\alpha) = 1 - F(-\alpha)$, and therefore $V_{I_2}(l) = 1 - F(\frac{I_2-I_1}{4} + \delta_l) = F(\frac{I_1-I_2}{4} - \delta_l)$. Consequently, if $\delta_l \leq 0$ and $z_l \in D(I_1, I_2)$, then $E_l = I_1^-$ implies that $V_{E_l} < V_{I_2}(l)$ and $P(W = E) = 0$.

For $\delta_l > 0$ and $z_l \in D(I_1, I_2)$, $V_{E_l} > V_{I_2}(l)$. Further, as $\varepsilon \to 0$, $V_{I_1}(l) \to F(\frac{I_2-I_1}{4} + \delta) - F(\frac{I_1-I_2}{4} + \delta) < F(\frac{I_2-I_1}{4}) - F(\frac{I_1-I_2}{4}) = 1 - F(\frac{I_1-I_2}{4}) - F(\frac{I_1-I_2}{4}) \leq \frac{1}{3}$. Therefore, $V_{E_l} > I_1(l), I_2(l)$ and successful entry is possible. By the symmetry of F, incumbents win districts in D and lose districts not in D, and the proof is complete.

It is obvious that I_1 wins districts with a median voter in $[\frac{3I_1+I_2}{4}, \frac{I_1+I_2}{2})$ and I_2 wins $(\frac{I_1+I_2}{2}, \frac{3I_1+I_2}{4}]$, with a tie in districts at $\frac{I_1+I_2}{2}$. Denote the intervals as the following $D(I_1) = [\frac{3I_1+I_2}{4}, \frac{I_1+I_2}{2}]$ and $D(I_2) = [\frac{I_1+I_2}{2}, \frac{3I_1+I_2}{4}]$.

Define $\overline{D}(I_1) = D(I_1) \cap [\underline{Z}, \overline{Z}]$, and likewise for $\overline{D}(I_2)$. These are the intervals of districts won by each incumbent that actually exist.

Define $M(I_1) = \int_{D(I_1)} g(z)dz$, and likewise for $M(I_2)$. These are the shares of the districts won by each incumbent.

1.6.2 Proof of Proposition 1.1

Define $H = (2\underline{Z}, 2\overline{Z})$, and $\hat{H} = closure(H)$. In this proof, and those to follow, I will repeatedly employ Lemma 1.1. It follows from the lemma that if $Z_i \in D(I_1, I_2)$ then $E_i = \emptyset$ if entrants are behaving rationally (as assumed), and if $Z_i \notin D(I_1, I_2)$ then $E_i \in \Re$ and the entrant is victorious. As the incumbent parties seek to primarily maximize their share of districts won, in most cases (Case 5 being the only exception) analyses of the intervals $D(I_1)$ and $D(I_2)$ are sufficient to show that an incumbent party has improved its situation via a deviation. Unless specified, I will attempt to show that for some deviation $\tilde{I}_j$, for some $j \in \{1,2\}$, the deviator's share of districts won must increase. For $j = 1$ this implies that $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}(I_1, I_2, E)$ for all $E' \in C_E^{\varepsilon}(\tilde{I}_1, I_2)$ and $E \in C_E^{\varepsilon}(I_1, I_2)$ for some ε, and, therefore, that $\{I_1, I_2\}$ can't constitute an equilibrium.

Case 1 $I_1, I_2 \in H$.

<u>*Case 1.1:*</u> $M(I_j) = 0$ for some $j \in \{1,2\}$

Suppose $\overline{D}(I_1) = M(I_1) = \emptyset$. If $I_2 \neq 0$ then set $\tilde{I}_1 = -I_2$. Applying Lemma 1.1, this implies $\overline{D}(I_1) = [\frac{3I_1+I_2}{4}, 0]$ and $M(\tilde{I}_1) > 0$. Therefore, $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$.

If $I_2 = 0$ then set $\tilde{I}_1 = \underline{Z}$. This implies $\overline{D}(I_1) = [\frac{3\underline{Z}}{4}, \frac{\underline{Z}}{2}]$ and $M(\tilde{I}_1) > 0$. Therefore, $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$.

Case 1.2: $M(I_1), M(I_2) > 0$.

This requires $\frac{I_1+I_2}{2} \in (\underline{Z}, \overline{Z})$.

Case 1.2.1: Suppose that $\frac{3I_1+I_2}{4} < \underline{Z}$ or $\frac{I_1+3I_2}{4} > \overline{Z}$. Therefore, $D(I_1) = [\frac{3I_1+I_2}{4}, \frac{I_1+I_2}{2}] \Longrightarrow \overline{D}(I_1) = [\underline{Z}, \frac{I_1+I_2}{2}]$. Consider the deviation $\tilde{I}_1 = I_1 + \delta$, where δ is s.t. $\frac{3\tilde{I}_1+I_2}{4} = \underline{Z}$. This implies that $D(\tilde{I}_1) = [\frac{3\tilde{I}_1+I_2}{4}, \frac{\tilde{I}_1+I_2}{2}] \Longrightarrow \overline{D}(I_1) = [\underline{Z}, \frac{\tilde{I}_1+I_2}{2}] = [\underline{Z}, \frac{I_1+I_2}{2} + \frac{\delta}{2}]$. Therefore, $\overline{D}(I_1) \subset \overline{D}(\tilde{I}_1)$, and $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$. Likewise for $\frac{I_1+3I_2}{4} > \overline{Z}$.

Case 1.2.2: $[\frac{3I_1+I_2}{4}, \frac{I_1+3I_2}{4}] = D = [\underline{Z}, \overline{Z}]$. This requires $I_1, I_2 \notin H$. See Case 4.

Case 1.2.3: $[\frac{3I_1+I_2}{4}, \frac{I_1+3I_2}{4}] = D \subset [\underline{Z}, \overline{Z}]$. Suppose that $|I_1| \geq |I_2|$. Consider the deviation $\tilde{I}_1 = I_1 - \gamma$, where $\gamma > 0$ and such that $\tilde{I}_1 \in H$. Thus, $D(\tilde{I}_1) = \overline{D}(\tilde{I}_1) = [\frac{3\tilde{I}_1+I_2}{4}, \frac{\tilde{I}_1+I_2}{2}] = [\frac{3I_1+I_2}{4} - \frac{3\gamma}{4}, \frac{I_1+I_2}{2} - \frac{\gamma}{2}]$, which implies the following relationship, $\overline{D}(\tilde{I}_1) = \overline{D}(I_1) - [\frac{I_1+I_2}{2} - \frac{\gamma}{2}, \frac{I_1+I_2}{2}] + [\frac{3I_1+I_2}{4} - \frac{3\gamma}{4}, \frac{3I_1+I_2}{4}]$. If g is strictly quasi-concave then, by Condition 1, $M(\tilde{I}_1) > M(I_1)$. If g is quasi-convex then for small enough γ, $\frac{3I_1+I_2}{4} < \frac{I_1+I_2}{2} - \frac{\gamma}{2} < 0$, and $M(\tilde{I}_1) > M(I_1)$ (as there is more density at the extremes). Therefore, for all g, $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$.

Case 2 $I_1, I_2 \in Z'/\hat{H}$.

Case 2.1: $I_1, I_2 < 2\underline{Z}$.

Lemma 1.1 implies that $M(I_1) = M(I_2) = 0$. Consider the deviation $\tilde{I}_2 = -I_1$. Therefore, $D(I_2) = [0, \frac{\tilde{I}_2}{2}] \Longrightarrow \overline{D}(I_2) = [0, \overline{Z}]$ as $I_1 < 2\underline{Z}$. Consequently, $M(I_2) > 0$ and $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

Case 2.2: $I_1 < 2\underline{Z}, I_2 > -2\underline{Z}$.

Let $|I_1| \geq |I_2|$. Therefore, $D(I_1) = [\frac{3I_1+I_2}{4}, \frac{I_1+I_2}{2}] \Longrightarrow \overline{D}(I_1) = [\underline{Z}, \frac{I_1+I_2}{2}]$. Consider the deviation, $\tilde{I}_1 = I_1 + \delta$, where δ is s.t. $\tilde{I}_1 < 2\underline{Z}$. Consequently, $D(\tilde{I}_1) = [\frac{3\tilde{I}_1+I_2}{4}, \frac{\tilde{I}_1+I_2}{2}] = [\frac{3I_1+I_2}{4} + \frac{3\delta}{4}, \frac{I_1+I_2}{2} + \frac{\delta}{2}] \Longrightarrow \overline{D}(\tilde{I}_1) = [\underline{Z}, \frac{I_1+I_2}{2} + \frac{\delta}{2}] \supset \overline{D}(I_1)$ as $\left|\tilde{I}_1\right|, |I_2| > 2\overline{Z}$. Thus, $M(\tilde{I}_1) > M(I_1)$ and $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$.

Case 3 $I_j \in Z', I_k \notin Z'$ for $j, k \in \{1, 2\}$ and $j \neq k$.

WOLOG assume that $I_1 \in Z', I_2 \notin Z'$. I will use the D notation from Lemma 1.1, though calling it D' here because $I_2 \notin Z'$ means the lemma may not be applicable.

What can be seen is that some of the arguments from the lemma can be preserved: Case 2 still applies but Case 1 does not (there may be successful entry between the incumbents). This implies that $D(I_1) \subseteq D'(I_1)$ and $D(I_2) \subseteq D'(I_2)$. That is, the set of districts that would be won (if they existed) is a subset of D'. WOLOG let $I_1 \leq 0$.

Case 3.1: $I_2 < 2Z^*$.

In this case $I_2 < I_1$, and therefore $D'(I_2) = [\frac{3I_2+I_1}{4}, \frac{I_2+I_1}{2}]$. As $\frac{I_2+I_1}{2} < Z^*, \overline{D}(I_2) = 0$. If $I_1 < 0$, the deviation of $\tilde{I}_2 = -I_1 \implies D(\tilde{I}_2) = [0, \frac{-I_1}{2}]$ (as now $\tilde{I}_2 \in Z'$) $\implies M(\tilde{I}_2) > 0$. Therefore, $O_{I_1}\left(\tilde{I}_1, I_2, E'\right) \succ O_{I_1}$. If $I_1 = 0$, the deviation of $\tilde{I}_2 = \overline{Z} \implies D(\tilde{I}_2) = [\frac{\overline{Z}}{2}, \frac{3\overline{Z}}{4}]$ (as now $\tilde{I}_2 \in Z'$) $\implies M(\tilde{I}_2) > 0$. Again, this implies $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$.

Case 3.2: $I_2 > -2Z^*$.

$D(I_2) \subseteq D'(I_2) = [\frac{I_1+I_2}{2}, \frac{I_1+3I_2}{4}] \implies \overline{D}(I_2) \subseteq \overline{D}'(I_2) = [\frac{I_1+I_2}{2}, \overline{Z}]$. If this set is empty then consider the deviations in Case 3.1 above. So suppose that $\overline{D}'(I_2) \neq \emptyset$ (this requires $\frac{I_1+I_2}{2} < \overline{Z}$) and consider the deviation $\tilde{I}_2 = -2Z^* - \varepsilon$ for $\varepsilon > 0$. Therefore, $D(\tilde{I}_2) = [\frac{I_1+\tilde{I}_2}{2}, \frac{I_1+3\tilde{I}_2}{4}] \implies \overline{D}(I_2) = [\frac{I_1+\tilde{I}_2}{2}, \overline{Z}]$ for small enough ε as $I_1 > 2Z^*$ (the lemma is now applicable). As $\tilde{I}_2 < I_2$ and $\frac{I_1+\tilde{I}_2}{2} > 0$, then $\overline{D}'(I_2) \subset \overline{D}(\tilde{I}_2)$ and it must be that $M(\tilde{I}_2) > M(I_2)$. Thus, $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

Case 4 $I_1 = 2\underline{Z}$ or $I_2 = 2\overline{Z}$.

Both of these locations are part of the equilibrium of the proposition. WOLOG suppose that $I_1 = 2\underline{Z}$. I will show that $I_2 = -2\underline{Z}$ is a strict best response for I_2. This proves existence of the equilibrium in the proposition, as well as ruling out any other equilibria involving either of these policy positions.

Case 4.1: $I_2 = -2\underline{Z}$.

$D(I_1) = [\underline{Z}, 0], D(I_2) = [0, -\underline{Z}] \implies M(I_2) = \frac{1}{2}$.

Case 4.2: $I_2 \leq \underline{Z}$.

As $I_1, I_2 \leq \underline{Z}$ then $\overline{D}'(I_2) = \emptyset$ and so $M(I_2) = 0$.

Case 4.3: $\underline{Z} < I_2 < -2\underline{Z}$.

Suppose that $M(I_2) > 0$ and consider $\tilde{I}_2 = I_2 + \alpha$, where $\alpha > 0$ such that $\tilde{I}_2 \leq 2\overline{Z}$. After the deviation I_2 will win additional districts with measure $\frac{3\alpha}{4}$ but lose districts with, at most, measure $\frac{\alpha}{2}$. If g is strictly quasi-concave then by Condition 1, $M(\tilde{I}_2) > M(I_2)$. So as $\tilde{I}_2 \to 2\overline{Z}$, $M(\tilde{I}_2)$ is increasing and approaching $M(I_2 = 2\overline{Z}) = \frac{1}{2}$. Therefore, for $I_2 < 2\overline{Z}$ it must be that $M(I_2) < \frac{1}{2}$. If g is strictly quasi-convex then this argument does not necessarily apply. If I_2 was to win districts in an interval of length $|\underline{Z}|$ on g, then it can be seen that the optimal location of this interval is $[\underline{Z}, 0]$, or $[0, \overline{Z}]$, in which case $M(I_2) = \frac{1}{2}$. For $I_2 < 2\overline{Z}$ the measure of $\overline{D}(I_2)$ is less than $|\underline{Z}|$. As G is strictly increasing once $G(.) > 0$, it must be that $M(I_2) < \frac{1}{2}$ for $I_2 < 2\overline{Z}$.

Case 4.4: $I_2 > -2\underline{Z}$.

$D'(I_2) = [\frac{2\underline{Z}+I_2}{2}, \frac{2\underline{Z}+3I_2}{4}] \Longrightarrow \overline{D}(I_2) \subseteq [\frac{2\underline{Z}+I_2}{2}, -\underline{Z}]$. As $2\underline{Z} + I_2 > 0$ it must be that $M(I_2) < \frac{1}{2}$.

Case 5 $I_1, I_2 \notin Z'$.

As mentioned at the beginning of the proof, for this case I will consider vote shares of the incumbent parties. This is necessary because there is entry in all districts for the incumbent locations I_1 and I_2, and the potential to deter entry in any district with a unilateral deviation cannot always be established. Therefore, to show that profitable deviations exist, I need look solely at total vote share as the share of districts won can only weakly increase. To establish this technique I will begin each subcase by establishing that $M(I_1|I_1, I_2) = M(I_2|I_1, I_2) = 0$.

Case 5.1: $I_1, I_2 \leq 2Z^*$ (or $\geq -2Z^*$).

Consider, for any district i, $E_i = [\max\{I_1, I_2\}] + \delta$ where $\delta > 0$. Obviously $\frac{\partial V_{E_i}}{\partial \delta} < 0$ and $V_{E_i} \to 1 - F(\max\{I_1, I_2\}) > \frac{1}{2}$ as $\delta \to 0^+$. Therefore, $E_i = [\max\{I_1, I_2\}]^+$ wins the district for the entrant and maximizes his vote share. Consequently, as this holds for all districts i, $M(I_1|I_1, I_2) = M(I_2|I_1, I_2) = 0$. Of course, if $F(\max\{I_1, I_2\}) = 0$ then $C^{\varepsilon}_{E_l}$ is an interval for all districts l and $\varepsilon > 0$.

Case 5.1.1: Suppose that $I_1 < I_2$. Then for district l with median $z_l \in [\underline{Z}, \overline{Z}]$, I_2's vote share is given by $V_{I_2}(l) \to F(I_2 - z_l) - F(\frac{I_2+I_1}{2} - z_l)$. If $\tilde{I}_2 = 2Z^* + \gamma$ then the optimal location for the entrant, for γ small enough, is still $E_l = \tilde{I}_2^+$. Now $V_{I_2}(l|I_1, \tilde{I}_2) \to F(2Z^* + \gamma - z_l) - F(\frac{2Z^*+\gamma+I_1}{2} - z_l) > F(I_2 - z_l) - F(\frac{I_2+I_1}{2} - z_l)$ as

$I_2, \tilde{I}_2 < \underline{Z}$, $f(2Z^* + \gamma - z_l) > 0$, and by Assumption 1.5, $f'(x) \geq 0$ for all $x \leq 0$. Consequently, as this holds in all districts, $V_{I_2}(I_1, \tilde{I}_2) > V_{I_2}$ and $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

Case 5.1.2: Suppose that $I_1 = I_2$. Then for district l with median $z_l \in [\underline{Z}, \overline{Z}]$, I_1's vote share is given by $V_{I_1}(l) \to \frac{1}{2}F(I_1 - z_l)$. By the continuity of f and g, it must be that $\int_{\underline{Z}}^{\overline{Z}} F(x - z_l)\, dz_l$ is also continuous in x. Therefore, if $F(I_1 - z_l) \neq 0$ for any z_l, δ small enough can be found s.t. $\tilde{I}_1 = I_1 - \delta \Longrightarrow V_{I_1}(\tilde{I}_1, I_2) = \int_{\underline{Z}}^{\overline{Z}} F\left(I_1 - \frac{\delta}{2} - z_l\right) dz_l > \frac{1}{2}\int_{\underline{Z}}^{\overline{Z}} F(I_1 - z_l)\, dz_l$. Thus, I_1's total vote share increases and $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$. So instead suppose that $F(I_1) = 0$ for all districts l. Consider then the deviation $\tilde{I}_1 = 2Z^*$. As in case 5.1.1, this implies that $E_l = \tilde{I}_1^+$ and $V_{I_1}(l|\tilde{I}_1, I_2) > 0$ for some z_l. Thus, I_1's vote share strictly increases in some districts and weakly increases in all districts. Therefore, $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$.

<u>*Case 5.2:*</u> $I_1 < 2Z^*, I_2 > -2Z^*$.

Without loss of generality let $|I_1| \leq |I_2|$. For district l with median $z_l \in [\underline{Z}, \overline{Z}]$, $E_l \notin (I_1, I_2) \Rightarrow V_{E_l} < F(Z^*) = \frac{1}{3}$ and the entrant doesn't win. So consider instead $E_l \in (I_1, I_2)$. In this case $V_{E_l} = F(\frac{I_2+E_l}{2} - z_l) - F(\frac{I_1+E_l}{2} - z_l)$ and $\frac{\partial V_{E_l}}{\partial E_l}.2 = f(\frac{I_2+E_l}{2} - z_l) - f(\frac{I_1+E_l}{2} - z_l)$. By Assumptions 1.1-1.5 the function V_{E_l} is quasi-concave. Suppose an interior maximum exists. Therefore, for some E_l, $f(\frac{I_2+E_l}{2} - z_l) = f(\frac{I_1+E_l}{2} - z_l)$ and $V_{I_1} = V_{I_2} < F(Z^*) = \frac{1}{3}$ as $(\frac{I_2-I_1}{2}) > 2Z^*$ (where $(\frac{I_2-I_1}{2})$ is the length of the interval of voters won by the entrant). If an interior maximum doesn't exist then $E_l \in \{I_1^+, I_2^-\}$. If $f(\frac{I_2+E_l}{2} - z_l) < f(\frac{I_1+E_l}{2} - z_l)$ for all $E_l \in (I_1, I_2)$ then the maximum is at $E_l = I_1^+$ and $V_{I_2} < V_{I_1} < F(Z^*) = \frac{1}{3}$. Similarly, $V_{I_1} < V_{I_2} < F(Z^*) = \frac{1}{3}$ if the reverse inequality holds. Therefore, for all districts l, an entrant locates between the incumbents and is victorious.

If an interior maximum exists then it is given by $\tilde{E}_l = 2z_l - (\frac{I_1+I_2}{2})$. As $|I_1| \leq |I_2|$, $\tilde{E} \leq 2z_l < I_2$. Thus, for an interior maximum to not exist it must be that $\tilde{E}_l \leq I_1$. Simple algebra shows that this requires $z_l \leq \frac{3I_1+I_2}{4}$. Therefore, for all z_l at which an interior maximum doesn't exist it must be that $E_l = I_1^+$ (as in these districts $E_l = I_2^- \Rightarrow f(\frac{I_2+E_l}{2} - z_l) < f(\frac{I_1+E_l}{2} - z_l)$, a contradiction). These results imply that for all districts such that $z_l \in [\underline{Z}, \overline{Z}]$, $V_{I_1}(l) \geq V_{I_2}(l)$.

Now consider a deviation by I_2 to $\tilde{I}_2 = -2Z^* - \rho$. For small enough ρ Case 2 of Lemma 1.1 still applies and the entrant locates between the incumbents or stays out. Suppose the entrant enters. Then, as $|I_1| > \left|\tilde{I}_2\right|$, $V_{I_1}(l|I_1, \tilde{I}_2, E) \leq V_{I_2}(l|I_1, \tilde{I}_2, E)$ for all districts by the same arguments as above. As $|\frac{\tilde{I}_2 - I_1}{2}| < |\frac{I_2 - I_1}{2}|$ the vote share for the entrant in each district must decline. And so $V_{I_1}(l|I_1, \tilde{I}_2, E) + V_{I_2}(l|I_1, \tilde{I}_2, E) > V_{I_1}(l) + V_{I_2}(l)$ for every l. As $V_{I_1}(l) \geq V_{I_2}(l)$ and $V_{I_1}(l|I_1, \tilde{I}_2, E) \leq V_{I_2}(l|I_1, \tilde{I}_2, E)$ we must have that $V_{I_2}(l|I_1, \tilde{I}_2, E) > V_{I_2}(l)$ for every l. If the entrant stays out then I_2's vote share in every district is at least $V_{I_2}(l|I_1, \tilde{I}_2, E)$. Therefore, $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

Case 6 $I_j \in H, I_k \in Z'/\hat{H}$ for $j, k \in \{1, 2\}$ and $j \neq k$.

<u>*Case 6.1:*</u> $I_1, I_2 \leq 0$.

As $I_1 \leq I_2$ It must be that $I_2 \in (2\underline{Z}, 0], I_1 \in (2Z^*, 2\underline{Z})$. $D(I_1) = [\frac{3I_1 + I_2}{4}, \frac{I_1 + I_2}{2}] \Longrightarrow \overline{D}(I_1) = \emptyset$ as $\frac{I_1 + I_2}{2} < \underline{Z}$. Thus $M(I_1) = 0$. If $I_2 < 0$ then consider the deviation $\tilde{I}_1 = -I_2$. If $I_2 = 0$ then consider the deviation $\tilde{I}_1 = \overline{Z}$. Both deviations imply, as shown previously, that $M(\tilde{I}_1) > 0$. Therefore, $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_2}$.

<u>*Case 6.2:*</u> $I_1, I_2 \geq 0$. Proceeds analogously to Case 6.1.

<u>*Case 6.3:*</u> $I_1 < 0$ and $I_2 \in \left(2\overline{Z}, -2Z^*\right)$.

$D(I_2) = [\frac{I_1 + I_2}{2}, \frac{I_1 + 3I_2}{4}] \Longrightarrow \overline{D}(I_2) = [\frac{I_1 + I_2}{2}, \overline{Z}]$, as $\frac{I_1 + 3I_2}{4} > \overline{Z}$. Now consider the deviation, $\tilde{I}_2 = I_2 - \alpha, \alpha > 0$, such that $\frac{I_1 + 3\tilde{I}_2}{4} = \overline{Z}$. Then $D(\tilde{I}_2) = [\frac{I_1 + \tilde{I}_2}{2}, \frac{I_1 + 3\tilde{I}_2}{4}] \Longrightarrow \overline{D}(\tilde{I}_2) = [\frac{I_1 + I_2 - \alpha}{2}, \overline{Z}]$. And so $D(I_2) \subset D(\tilde{I}_2)$ and I_2 is strictly better off. Therefore, $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

1.6.3 Proof of Proposition 1.1A

The proof is identical to that for Proposition 1.1, with Condition 1A substituted for Condition 1. The extra restriction in Condition 1A ensures that the arguments of Proposition 1.1 can also be used to prove Proposition 1.1A.

1.6.4 Proof of Proposition 1.2

Using the notation of the previous section, we can see that for $I_1 \leq I_2$ we have $D' = [\frac{3I_1+I_2}{4}, \frac{I_1+3I_2}{4}]$. As $D \subseteq D'$ to preclude entry we require $\frac{3I_1+I_2}{4} \leq \underline{Z}$ and $\frac{I_1+3I_2}{4} \geq \overline{Z}$. Solving these two requirements simultaneously and recalling that $\underline{Z} = -\overline{Z}$, we have $I_1 \leq 2\underline{Z}$ and $I_2 \geq -2\underline{Z}$. As $\underline{Z} < Z^*$ this implies that $I_1 < 2Z^*$ and $I_2 > -2Z^*$, but then we will have entry between the incumbents in every district (see Proposition 1.1, case 5). So there does not exist a pair of locations for the incumbents which are able to preclude entry in every district. Thus, there does not exist an equilibrium which precludes entry in all districts.

1.6.5 Proof of Proposition 1.3

This proof is very similar to that of Proposition 1.1, in many instances the only difference being a change in the domain of a case. Define $H^\# = \left(2Z^\#, -2Z^\#\right)$, and $\hat{H}^\# = closure\left(H^\#\right)$.

Case 1 $I_1 I_2 \in H^\#$.

Proceed as with Case 1 in Proposition 1.1. As g violates Condition 1 the quasi-convexity requirement can be ignored.

Case 2 $I_1, I_2 \in Z'/\hat{H}^\#$.

<u>*Case 2.1:*</u> $2Z^\# \leq \underline{Z}$.

Case 2.1.1: $I_1, I_2 < 2Z^\#$. $E_i = I_2^+$ in all districts and implies $M(I_1) = M(I_2) = 0$. The deviation $\tilde{I}_2 = -I_1$ requires that $D(I_2) = [0, \frac{\tilde{I}_2}{2}] \Longrightarrow M(I_2) > 0$ and, therefore, $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

Case 2.1.2: $I_1 < 2Z^\#, I_2 > -2Z^\#$. Suppose that $|I_1| - \gamma = |I_2|$, for some $\gamma > 0$. Thus $D(I_1) = [\frac{-I_2}{2} - \frac{3\gamma}{4}, \frac{-\gamma}{2}]$ and $D(I_2) = [\frac{-\gamma}{2}, \frac{I_2}{2} - \frac{\gamma}{4}]$. If $\frac{3I_1+I_2}{4} < \underline{Z}$ then the deviation $\tilde{I}_1 = I_1 + \delta$, such that $\frac{3\tilde{I}_1+I_2}{4} = \underline{Z}$, implies that $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$ (see Proposition 1.1, case 2). Therefore, assume that $\frac{3I_1+I_2}{4} \geq \underline{Z}$ and consider the following two deviations: (1) $\tilde{I}_1 = -I_2$, and (2) $\tilde{I}_2 = -I_1$. For deviation (1), $D(\tilde{I}_1) = [\frac{-I_2}{2}, 0] \Longrightarrow$

$M(\tilde{I}_1) = M(I_1) + [G(0) - G(\frac{-\gamma}{2})] - [G(\frac{-I_2}{2}) - G(\frac{-I_2}{2} - \frac{3\gamma}{4})]$. If $[G(0) - G(\frac{-\gamma}{2})] > [G(\frac{-I_2}{2}) - G(\frac{-I_2}{2} - \frac{3\gamma}{4})]$ then $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$ and $\{I_1, I_2\}$ can't constitute an equilibrium. If this inequality doesn't hold then consider deviation (2). In this case $D(\tilde{I}_2) = [0, \frac{I_2+\gamma}{2}] \Longrightarrow M(\tilde{I}_2) = M(I_2) + [G(\frac{I_2+\gamma}{2}) - G(\frac{I_2}{2} - \frac{\gamma}{4})] - [G(0) - G(\frac{-\gamma}{2})]$. By the symmetry of g, $G(\frac{I_2+\gamma}{2}) - G(\frac{I_2}{2} - \frac{\gamma}{4}) > G(\frac{-I_2}{2}) - G(\frac{-I_2}{2} - \frac{3\gamma}{4})$. Thus, if deviation (1) isn't profitable for I_1 then $[G(\frac{I_2+\gamma}{2}) - G(\frac{I_2}{2} - \frac{\gamma}{4})] > [G(0) - G(\frac{-\gamma}{2})] \Longrightarrow M(\tilde{I}_2) > M(I_2)$ which implies that $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$ and deviation (2) is profitable for I_2.

Suppose instead that $|I_1| = |I_2|$, which implies $D(I_1) = [\frac{I_1}{2}, 0]$. Consider the deviation $\tilde{I}_1 = I_1 + \alpha$, where $\alpha > 0$, such that $D(\tilde{I}_1) = [\frac{I_1}{2} + \frac{3\alpha}{4}, \frac{\alpha}{2}]$. As $\frac{I_1}{2} < Z^{\#}$ then it must be that $g(\frac{I_1}{2}) < \frac{2}{3}g(0)$. Because g is continuous there exists an α small enough such that $\forall \alpha' < \alpha$, $g(\frac{I_1}{2} + \frac{3\alpha'}{4}) < \frac{2}{3}g(\frac{\alpha'}{2})$. This implies that $M(\tilde{I}_1) > M(I_1)$ and $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$. as I_1 gains more districts in the center than it loses on the fringe.

<u>*Case 2.2:* $\underline{Z} < 2Z^{\#}$</u>.

For $I_1 < 0, I_2 > 0$ this is the same as for $\underline{Z} \geq 2Z^{\#}$. So assume that $I_1, I_2 < 2Z^{\#}$. We recall that by setting $\tilde{I}_1 = -I_2$, $M(I_1) > 0$, and likewise for I_2, so equilibrium requires that $M(I_1), M(I_2) > 0$. Therefore, as $I_1 = I_2 \Longrightarrow M(I_1) = M(I_2) = 0$, in any equilibrium it must be that $I_1 < I_2 < 2Z^{\#}$. Consider the deviation $\tilde{I}_2 = I_2 + \delta$, such that $\tilde{I}_2 < 2Z^{\#}$. This implies $D(\tilde{I}_2) = D(I_2) + [\frac{I_1+3I_2}{4}, \frac{I_1+3I_2}{4} + \frac{3\delta}{4}] - [\frac{I_1+I_2}{2}, \frac{I_1+I_2}{2} + \frac{\delta}{2}]$. As $\tilde{I}_2 < 0$ and $M(I_2) > 0$, g is strictly increasing on this domain which implies that $M(\tilde{I}_2) > M(I_2)$. Thus $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

Case 3 $I_j \in Z', I_k \notin Z'$ for $j, k \in \{1, 2\}$ and $j \neq k$.

Proceed as with Case 3 in Proposition 1.1.

Case 4 $I_1 = 2Z^{\#}$ or $I_2 = -2Z^{\#}$.

Both of these locations are part of the potential equilibrium of the proposition. WOLOG suppose that $I_1 = 2Z^{\#}$. I will show that $I_2 = -2Z^{\#}$ is a strict best response for I_2, given that g is concave and show that when g is not concave $I_2 \neq -2Z^{\#}$ cannot constitute an equilibrium.

If $I_2 = -2Z^\#$ then $D(I_1) = [Z^\#, 0], D(I_2) = [0, -Z^\#] \Longrightarrow M(I_2) = \frac{1}{2} - G(Z^\#)$. Using the same techniques as in Proposition 1.1, all I_2 such that $I_2 \geq I_1$ can be shown to imply that $M(I_2) < \frac{1}{2} - G(Z^\#)$ and thus can't constitute equilibria. As Condition 1 I also must consider the case of $I_1 < I_2$ when $2Z^\# > \underline{Z}$ (when there are districts to the left of I_1).

Suppose that g is concave. I will show that there can't exist more density on the flanks that would entice I_2 to deviate. As $g(Z^\#) = \frac{2}{3}g(0)$, the concavity of g implies that $g(2Z^\#) < \frac{1}{3}g(0)$ and $g(3Z^\#) = 0$. Therefore, for $I_2 < I_1$, $\overline{D}(I_2) \subset [I_2, I_1]$ and $|\overline{D}(I_2)| \leq \frac{|Z^\#|}{4}$. Thus, $M(I_2)$ is bounded by $G(2Z^\#) - G(2Z^\# - \frac{|Z^\#|}{4})$ which by the concavity of g implies $M(I_2) < \frac{1}{2} - G(Z^\#)$. Consequently, $\{I_1, I_2\} = \{2Z^\#, -2Z^\#\}$ is a strict equilibrium, and therefore the only equilibrium involving $I_1 = 2Z^\#$ or $I_2 = -2Z^\#$.

Suppose instead that g is not concave. Then it is possible that I_2 may wish to locate at such that $I_2 < I_1$. If this is the case then consider the incentives of I_1. If $\tilde{I}_1 = I_1 + \delta$, then $\overline{D}(\tilde{I}_1) = [\frac{\tilde{I}_1+I_2}{2}, \frac{3\tilde{I}_1+I_2}{4}] = \overline{D}(I) + [\frac{3I_1+I_2}{4}, \frac{3I_1+I_2}{4} + \frac{3}{4}\delta] - [\frac{I_1+I_2}{2}, \frac{I_1+I_2}{2} + \frac{1}{2}\delta]$. As, for small enough δ, g is strictly increasing over this range it must be that $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$. Thus, for quasi-concave g functions $I_2 \neq -2Z^\#$ can't support an equilibrium and if an equilibrium exists in this domain it must be $\{I_1, I_2\} = \{2Z^\#, -2Z^\#\}$.

Case 5 $I_1, I_2 \notin Z'$.

Proceed as with Case 5 in Proposition 1.1.

Case 6 $I_j \in H^\#, I_k \in Z'/\hat{H}^\#$ for $j, k \in \{1, 2\}$ and $j \neq k$.

<u>Case 6.1:</u> $I_1, I_2 \leq 0$.

Therefore, $I_1 \in [2Z^*, 2Z^\#)$ and $I_2 \in (2Z^\#, 0]$.

Case 6.1.1: $\frac{I_1+I_2}{2} \leq \underline{Z}$. Proceed as in Case 6 of Proposition 1.1.

Case 6.1.2: $Z^\# > \frac{I_1+I_2}{2} > \underline{Z}$. Thus $D(I_2) = \overline{D}(I_2) = [\frac{I_1+I_2}{2}, \frac{I_1+3I_2}{4}]$. Suppose $I_2 < 0$ and consider the deviation $\tilde{I}_2 = \frac{I_2}{2}$. Then $D(\tilde{I}_2) = \overline{D}(\tilde{I}_2) = [\frac{I_1+\tilde{I}_2}{2}, \frac{I_1+3\tilde{I}_2}{4}] = [\frac{I_1+I_2}{2} - \frac{I_2}{4}, \frac{I_1+3I_2}{4} - \frac{3I_2}{8}]$. As $\frac{I_2}{2} < 0$ then the strict quasi-concavity and symmetry of g

imply that $M(I_2) < M(\tilde{I}_2)$, and $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$. So suppose instead that $I_2 = 0$ and consider the deviation $\tilde{I}_2 = \alpha$, where $\alpha > 0$, such that $|\frac{I_1+3\alpha}{4}| < 0$. This implies that $D(\tilde{I}_2) = \overline{D}(\tilde{I}_2) = [\frac{I_1+\tilde{I}_2}{2}, \frac{I_1+3\tilde{I}_2}{4}] = [\frac{I_1+I_2}{2} + \frac{\alpha}{2}, \frac{I_1+3I_2}{4} + \frac{3\alpha}{4}]$, and once again by the strict quasi-concavity and symmetry of g, $M(I_2) < M(\tilde{I}_2)$, and $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

Case 6.2: $I_1, I_2 \geq 0$

This case is analogous to Case 6.1.

Case 6.3: $I_1 \in (2Z^{\#}, 0)$ and $I_2 \in (-2Z^{\#}, -2Z^{*}]$.

Case 6.3.1: $\frac{I_1+3I_2}{4} > \underline{Z}$. Consider the deviation $\tilde{I}_2 = I_2 - \rho$, such that $\frac{I_1+3\tilde{I}_2}{4} = \underline{Z}$ (as in Proposition 1.1). Therefore, $\overline{D}(I_1) \subset \overline{D}(\tilde{I}_1)$ and $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$.

Case 6.3.2: $\frac{I_1+3I_2}{4} \leq \underline{Z}$. Consider dual deviations by the incumbents as done in Case 2 of this result. At least one incumbent has a positive incentive to deviate and, therefore, $\{I_1, I_2\}$ can't constitute an equilibrium.

1.6.6 Proof of Theorem 1.1

This result is merely a combination of the previous results. It is stated in order to give a clear representation of what has been established. Its proof simply refers to the previous results.

Sufficient$\Longrightarrow$ Conditions 1 and 2 satisfy the requirements of Proposition 1.1. No additional entry occurs and the final outcome involves two parties. Duverger's Law holds.

Necessary$\Longrightarrow$ If Condition 2 is violated then the requirements for Proposition 1.2 are satisfied. Third party entry occurs in every district and the final outcome involves three or more parties. Duverger's Law fails. If Condition 1 is violated then the requirements for Proposition 1.3 are satisfied. Third party entry occurs in intervals of districts on the edges of the distribution of district median voters. The final outcome involves three or more parties and Duverger's Law fails.

Chapter 2 Electoral Competition and the Run-Off Rule

Abstract

Despite the wide range of voting rules in use in the world's electoral systems, most academic research has focused on only a few of these. One rule that has received insufficient attention is the frequently used run-off rule. In this chapter, I describe the incentives faced by candidates and voters in a model of electoral competition with entry under the run-off rule and characterize the equilibria. I find that a continuum of equilibria exist in which one of two incumbent parties always win. This result is found to be robust to variations in the motivations of the entrant, the timing of entry decisions, as well as the preferences of the parties. Significantly, this implies that if parties are victory seeking then only two parties will enter the election. This result is then reconciled with what Riker (1982) has called "Duverger's Hypothesis" and a more precise formulation is proposed. I also consider an extension of the model to simultaneous competition for multiple districts and show that the results are robust to a limited amount of heterogeneity across districts.

2.1 Introduction

Over time there has arisen a substantial body of theoretical work that has explored the intricacies of electoral competition. However, most of this work has focused on the plurality rule and, to a lesser extent, proportional representation. To fully understand electoral incentives, political parties, and the choices they make, this analysis must be extended to a broader set of environments. This chapter aims to continue progress in this direction by developing a model of electoral competition with entry under the run-off rule.[1]

The run-off rule is a repeated process, where at each stage one or more parties are eliminated. The winner is, naturally enough, the final remaining party. The principle feature of this process is that, unlike competition under the plurality rule, it may not be in the interests of each party to maximize its vote share in the first round. As a result it has long been conjectured that the run-off rule does not provide the incentives for a two-party system to develop. This prediction, which groups the run-off rule with proportional representation, goes at least as far back as Lowell (1896) and Holcombe (1910). However, as is also the case with the plurality rule, these ideas are now most closely associated with Duverger (1954) and are referred to by Riker (1982) as "Duverger's Hypothesis."

To support this conjecture, these and other authors (e.g., Wright and Riker (1989), Shugart and Taagepera (1994)) have documented extensive empirical support.[2] This support, however, is not overwhelming and left Riker (1982, p. 760) to conclude that "we can therefore abandon Duverger's hypothesis in its deterministic form."[3] The

[1]Other rules that have received some attention include the alternative vote, scoring rules, and Condorcet procedures (see, for example, Brams and Fishburn (1983), Cox (1985, 1987), Myerson (1993), and Myerson and Weber (1993)). The existing work on the run-off rule will be discussed in Section 2.4.2.

[2]The literature has always grouped data from primary and legislative elections together when conducting analysis. However, it is not clear that such a combination is as innocuous as it first appears. Studies of primary elections (see Bartels (1988)) have shown that these intra-party contests are typically not fought on policy dimensions and instead are fought on candidate quality comparisons. This is in contrast to legislative elections that place more weight on policy comparisons. As it is policy driven contests that are analyzed here the evidence provided by Shugart and Taagepera (who study presidential elections) is perhaps more appropriate than the evidence of Wright and Riker (who study US gubernatorial primaries).

[3]For the case of legislative elections, Duverger (1954) presents the case of Australian parlia-

open question, therefore, is why, under some circumstances, two parties can deter further entry when competition is under the run-off rule.

In this chapter I attempt to provide an explanation to this very question. I show that, even under the run-off rule, two parties can choose platforms that will deter additional successful entry if the initial conditions of the electoral contest allow them to establish their positions. Further, I prove that not only is successful entry deterrence possible, but also that it is supportable as equilibrium behavior.

To establish these results I study a model with three parties and sincere voting. The parties, or candidates, are completely office motivated (i.e., they are Downsian) and are free to choose their policy platforms. I characterize the set of equilibria to the following dynamic game: firstly, two incumbent parties choose locations, followed by a potential entrant, and then all entered parties compete in the election. This ordering was employed to enable comparison with the results of Palfrey (1984) for the plurality rule, though the results presented here do not depend critically on this timing scheme. I study two cases for this model: when the entrant must enter, and when the entrant only enters if it has a positive probability of victory.

Solving this model, I find that there exist equilibria in which one of the two incumbent parties must win with certainty. In fact, I find that in all equilibria the entrant loses with certainty, and therefore doesn't enter when it has the choice. This final possibility implies a two-party outcome[4] as the incumbent parties are able to deter subsequent entry. This equilibrium correlates well with the two-party outcome under the run-off rule in Australia and the other counterexamples to "Duverger's Hypothesis."[5]

Further, and most surprisingly, I find that the location choices of the incumbent

mentary elections as an unexplained counterexample. Shugart and Taagepera (1994) calculate the effective number of candidates in presidential run-off elections and find many instances in which it is less than three. With respect to primary elections, several counterexamples are documented by Canon (1978) and Wright and Riker (1989). The latter authors find that, for southern US gubernatorial elections between 1956 and 1982, the number of run-off elections in which there are only two effective candidates (over 5% of the vote) is over $\frac{1}{4}$ the number of similar contests under the plurality rule.

[4]In that only the two incumbent parties enter and have a chance of victory.

[5]Though there are three major parties in Australia, one of these parties, the National Party, is effectively the rural arm of the conservative Liberal Party (see Riker (1982)).

parties are unaffected by whichever of the two entry assumptions is made. Under both assumptions there exists a continuum of equilibria in which the incumbent parties choose symmetric positions about the median voter. There is also an equilibrium in which both incumbent parties locate at the median.

These results serve to provide an explanation of when and why "Duverger's Hypothesis" will fail. However, even though the two-party outcomes found here permit the hypothesis to be rejected as a deterministic theory, a closer reading of Duverger shows that they are not, in fact, at odds with his original intuition. This is because the two results, both here and by Duverger, incorporate differing assumptions, either explicit or implicit, about the timing of the entry decision.

The results here suggest that if there are currently two competing parties then no more will enter. In contrast, the intuition of Duverger implies that if there are more than two parties already competing then these parties will not necessarily have the incentive to exit the election. These arguments are not mutually exclusive, and the multiplicity of real world outcomes under the run-off rule support such a conclusion. To deal with these subtleties of timing I present a more general statement of "Duverger's Hypothesis." Perhaps most significantly, these results imply that outcomes under the run-off rule are sensitive to the political environment in which they are introduced.

The discussion presented so far has considered only a generic "run-off rule." However, there are, in fact, many variants of the run-off rule. The most basic line of division is whether the repeated rounds involve repeated voting or merely repeated counting.[6] If voting is repeated then at each round voters are required to cast a new ballot to determine which parties shall progress. The most common form is the "dual-ballot," though obviously there could be any number of ballots. In the dual-ballot system all parties stand in the first round and voters cast a single, nontransferable vote.[7] Unless one party gains a majority, in which case it is declared the winner immediately, the two leading parties then progress to the second round and all other

[6] Riker (1982).

[7] Using the terminology of Cox (1997) throughout.

parties are eliminated.[8] The winner is decided from these remaining parties by a simple majority vote. This system is currently used for presidential elections in France and several Latin American countries, and was used in Israel between 1996 and 2001 to elect a prime minister.

If counting, rather than voting, is repeated then voters cast only one ballot. Instead of indicating only a most preferred party, the voters will rank the parties in some way. A common form of this rule is the "alternative vote," which is used for lower house state and federal elections in Australia.[9] With this rule voters are required to list their preferences for all parties from first to last. The number of first preferences are tallied for each party and the one with the fewest is eliminated. The ballots for this party are then redistributed according to the second preference listed. These votes are again tallied and the party with the lowest total is eliminated. This process is repeated until there is only one remaining party, which is then declared the winner. With this counting rule the number of rounds must be one less than the number of parties. Of course, different counting rules can be developed that produce any number of rounds.

The model developed here is consistent with both the alternative vote rule and the dual-ballot rule. Thus, the results obtained are applicable to both. This consistency is the product of three assumptions: three parties, full information and sincere voting. With only three parties, the alternative vote can have at most two rounds, as with a dual-ballot. With full information and sincere voting nothing can change in the interval between the two ballots and, consequently, the dual ballots could be conducted simultaneously.[10]

However, this equivalence may only be superficial. One difference that might arise between the alternative vote rule and the dual-ballot rule is the motivations of the parties. A natural primary objective, common to both rules, is that parties seek to maximize their probability of victory. However, the appropriate secondary objective

[8] A further variant allows any parties to progress to the second round, making the first round purely demonstrative.

[9] With the exception of the state of Tasmania (Wright (1986)).

[10] The implied assumption that parties cannot change their platforms between ballots is also required here. The appropriateness of this assumption will be discussed more on page 70.

may differ between the two rules. With a dual-ballot parties may want to maximize their chances of making the run-off stage in order to maximize their exposure by competing on an additional election day. In contrast, under the alternative vote rule parties may want to maximize their primary vote share, which is their vote share in the first round, as this maximizes their exposure (and possibly even their public funding). Interestingly, I find that the equilibria are unaffected by whichever of these assumptions on preferences is made.

Finally, I also explore a recent extension, introduced for the plurality rule in Chapter 1, to consider simultaneous competition for multiple districts (in which the parties enter candidates in each district). I find that the equilibria of the single district case are robust to some district heterogeneity. However, if the heterogeneity is too extreme then third party entry is inevitable. To make this precise I characterize the sufficient and necessary conditions such that two-party equilibria exist.

Despite the concentration of theoretical work on only a few voting rules, there are several papers that consider the run-off rule to varying degrees. Osborne and Slivinski (1996), however, is the only one, at least to the best of my knowledge, that considers a spatial model of candidate competition under the run-off rule.[11] Osborne and Slivinski construct a model of citizen-candidates under both the plurality and run-off rules. The primary difference of their model is that they assume candidates are policy oriented and thus, most importantly, policy restricted. They consider the opposite extreme to that modelled here by assuming candidates must choose their true ideal point as their campaign platform. Consequently, the intuition of their results is significantly different to the intuition presented here. Though, even at this opposite extreme, they find that for certain parameter values (of cost of entry and benefit of office) two-party equilibria exist. The connection between this paper and the current work will be discussed in more detail in Section 2.4.2.

The remainder of the chapter is organized as follows. Section 2 contains the single district model and Section 3 the results. In Section 4, I discuss the features of the

[11]Other related papers include Greenberg and Shepsle (1987), and Myerson (1993). These papers will be discussed on page 70.

model and results, and how they relate to "Duverger's Hypothesis." In Section 5, I introduce the extension to simultaneous competition for multiple districts and present the results for this environment. Section 6 concludes.

2.2 The Model

The model is one of electoral competition with entry in a single district. The participants and order of play are as follows. There are two incumbent parties who choose their platforms simultaneously. A potential entrant then makes an entry decision, and if it chooses to enter it selects a platform position. The entered parties then engage in the election. This ordering was chosen to enable comparison with the results of Palfrey (1984) and Callander (1999) for the plurality rule, though it should be noted that the results presented here do not depend critically on this particular timing scheme.[12] The comparison of the results presented here and those for the plurality rule, along with a discussion of the robustness of the results to the timing scheme and other factors, will be performed in Section 2.4. I will denote the two incumbents by I_1 and I_2, and the entrant by E.

I will present the equilibria under two different assumptions on entrant behavior. These assumptions revolve around whether the party would enter even if it was certain it was going to lose. In the first treatment I will assume that one entrant (and only one entrant) will always enter, even if its probability of winning is zero. This is the assumption used by Palfrey (1984) in analyzing the plurality rule. The second treatment will assume that the potential entrant will enter only if it has a strictly positive probability of victory.[13] This decision rule captures the belief that political parties are designed to win elections. It will be seen that these two alternative assumptions, despite their divergent behavioral requirements, lead to the same equilibrium location choices by the two incumbent parties.

[12]Palfrey (1984) does not allow the entrant to stay out of the election, regardless of whether it can win, and therefore always produces three party outcomes. Callander (1999) extends this model and allows third parties to not enter if their probability of victory is zero.

[13]This extension for the plurality rule was examined in Callander (1999) and Chapter 1.

The issue space is the real line, $\Re$. There is a continuum of voters with symmetric, single peaked preferences over the issue space. The voters' ideal points are distributed according to the nondegenerate cdf F and the corresponding pdf, f. These functions have the following properties.

Assumption 2.1 *For all* $\alpha < 0$ *for which* $F(\alpha) > 0$*, the function* F *is strictly increasing on* $(\alpha, -\alpha)$.

Assumption 2.2 F *is continuous and twice differentiable at all points* $x \in \Re$ *such that* $F(x) \in (0, 1)$.[14]

Assumption 2.3 $F(x) = 1 - F(x)\ \forall x \in \Re$.

Assumption 2.4 $f'(x) \geq 0\ \forall x \leq 0,$ *and* $f'(x) \leq 0\ \forall x \geq 0$.

These assumptions specify that the distribution of voters' ideal points is symmetric about zero, and that the mass at any point is at least as great as at any point further from zero. This requires f to be quasi-concave. It can be seen that the uniform distribution satisfies these conditions. Assumption 2.1 ensures that there are no gaps in the distribution but without assuming that voter ideal points span all of $\Re$ (that is, voter ideal points can be contained in a bounded interval, for example [-1,1]).

Voters are assumed to be sincere and so vote for the party closest to their ideal point. I will further assume that if a voter is indifferent between the two incumbents then they randomize, but if they are indifferent between an incumbent and the entrant then they vote for the incumbent.[15,16] This assumption, stated formally below, prevents entrants from wanting to locate on top of an incumbent. It is a formalization of the notion that voters have a preference for established parties if all else is

[14] This specification allows for discontinuities in F at only two points: the boundary points of the support of f. This possibility permits the uniform, among others, as a possible distribution.

[15] This assumption is not typical of the plurality rule literature, though the plurality rule results of Callander (1999) and Chapter 1 do not change if this assumption is added. However, it is crucial to the run-off results, as otherwise entry prevention would not be possible (the entrant could locate on top of either incumbent and obtain a positive probability of victory).

[16] Alternatively, it could be assumed that ties in the overall election between an incumbent and the entrant are decided in favor of the incumbent, and that ties between incumbents are decided randomly.

the same. Any ties in the election are then decided randomly. Denote voter i's ideal point v_i and her vote choice by $vote\,(i)$. In an abuse of notation, let I_1, I_2 and E represent the parties electoral platforms.

Assumption 2.5 *If* $|v_i - E| < |v_i - I_j|$ *for* $j = 1, 2$ *then* $vote(i) = E$*. Otherwise,* $vote(i) = I_j$ *if* $|v_i - I_j| < |v_i - I_k|$ *where* $j, k = 1, 2$ *and* $j \neq k$*. If* $|v_i - I_1| = |v_i - I_2|$ *then* $prob[vote(i) = I_j] = \frac{1}{2}$ *for* $j = 1, 2$*.*

It should be noted here that this assumption does not place any restrictions on the voter's utility function other than that utility is decreasing in the distance from her ideal point. More specifically, a quadratic loss utility function is allowable with this assumption.

The run-off rule is used to map votes into an outcome. The run-off rule is one of elimination. Once a point is reached such that there are only two competing parties, then the one with the greatest vote share is the winner.[17] If the entrant does not enter then the incumbents compete by this rule. However, if the entrant chooses to contest the election then there are two stages. After the first stage, the party with the lowest vote share is eliminated and the remaining two parties compete as above. The winner of the election will be denoted by W.

Parties are free to locate at any point in the policy space, $\Re$, but must maintain this point for all rounds of the election. The parties have three-dimensional lexicographic preferences. The primary dimension is probability of victory, the second dimension is probability of surviving until the second round of the election, and the third dimension is primary vote share, which is the proportion of voters whose first preference is that party.[18]

[17]Thus, for a two-party contest this is equivalent to plurality rule. This obviously implies that if the entrant did not exist then this model is identical to one in which the plurality rule is used to count votes. As such, I will always assume that the entrant exists.

[18]The plurality results of Callander (1999) and Chapter 1, which employ similar preferences, would not change if instead it was assumed that parties simply vote maximize. This was the approach of Palfrey (1984). However, when considering the entrant's decision under the run-off rule vote maximization and the maximization of probability of victory do not necessarily coincide. As the probability of victory dictates the entry decision of this party it would seem natural to assume that this rule also dictates the location decision.

The most appropriate specification of preferences would seem to depend on whether the model is one of a single ballot alternative vote rule, or whether it is a dual-ballot process. If the latter, then it would seem sensible that the second dimension of preferences is the probability of making the second round. By making the second round a party would maximize its exposure as it is competing on an additional election day. Further, such an achievement confirms the party as the main opponent of the victor.

However, in a single ballot alternative vote election this justification does not exist. Rather, it would seem appropriate for a party to maximize its primary vote share. This is because, at least in Australia, the results from alternative vote elections report primary vote levels and which party is the winner. They do not report which parties survived the rounds. Indeed, public funding of political candidates is a function of primary vote share. Thus, maximum exposure would most likely come from maximizing primary vote share. These preferences are consistent with the specification above, and are simply the special case in which the second dimension of preferences is given zero weight.

Significantly, the results of the model are unaffected by whichever of these two specifications is assumed. Due to the incumbents difficulty in preventing successful entry, the same intuition holds whether or not the second dimension is given positive weight. Therefore, this generality of preferences does not prohibit the model from being an accurate representation of both the single ballot alternative vote rule and the dual-ballot rule.

Lexicographic preferences dictate that if a party has a set of points that maximize its probability of victory it chooses the point in this set that maximizes its probability of reaching the second round. And if this set is not a singleton then it chooses a point that maximizes its vote share. I assume that if there is more than one point in this set that maximizes a party's vote share then the party randomizes equally over these points. Define $P(j)$, $Q(j)$, and V_j, $j = I_1, I_2, E$, to be, respectively, the probability of victory, probability of making the second round, and vote share, of party j. The outcome function for party $j \in \{I_1, I_2, E\}$ can be written as follows.

$$O_j(I_1, I_2, E) = (P(j|I_1, I_2, E), Q(j|I_1, I_2, E), V_j\,(I_1, I_2, E))$$

Strict (weak) preference for one outcome over another is denoted by the binary relation $\succ$ ($\succeq$), where $A \succ B$ represents the situation in which outcome A is strictly preferred to outcome B.

Given the positions of the incumbents there may not exist an optimal location choice for the entrant. This technicality arises when the entrant attempts to maximize its vote share over the set of points that maximize its probability of winning and of making the second round. The probabilities of winning and making the second round for E can only take on finite sets of values (as there are only three parties and voting is deterministic) and so a set of maximizers over these dimensions can always be found.

A variant of the limit equilibrium concept introduced by Palfrey (1984) is employed to deal with this problem. If a vote maximizing point doesn't exist then the entrant 'almost' maximizes its vote share when choosing from the set of points which maximize its probability of winning and making the second round. A perturbed game is defined for each ε, where ε is how close each E comes to maximizing its vote share. An equilibrium is defined as any pair of strategies for I_1 and I_2 which are best responses to each other for an infinite sequence of the perturbed games, with the perturbation approaching zero in the limit.

The difference between this equilibrium and that of Palfrey is that his parties simply maximize vote share and their preferences are representable by a utility function. The parties modelled here have lexicographic preferences and such a general representation is not possible. Consequently the definition of equilibrium must remain in terms of primitive preferences. Technically the definition introduced here is stricter than Palfrey's, though the results are unaffected by this additional requirement.[19,20]

The set of points that maximize E's probability of victory is defined as follows.

$$X(I_1, I_2) = \arg\max_{x \in \Re}\{P(E|I_1, I_2, x)\}$$

From X the set of points that maximize E's probability of making the second round is given by X'.

$$X'(I_1, I_2) = \arg \max_{x \in X(I_1, I_2)} \{Q\left(E|I_1, I_2, x\right)\}$$

If the parties do not care about making the second round, and so the second dimension receives zero weight, then set $X'\left(I_1, I_2\right) = X\left(I_1, I_2\right)$.

Obviously, if the entrant chooses to not enter the election then $E = \emptyset$. If E does enter then the set of points that it equally randomizes over, for a given ε, is given by C_E^ε, where,

$$C_E^\varepsilon(I_1, I_2) = \{E \in X'(I_1, I_2) | V_E(I_1, I_2, E) > V_E(I_1, I_2, y) - \varepsilon, \forall y \in X'(I_1, I_2)\}$$

Anticipating this entry decision the expected outcome for the incumbents, given their own locations, is the expectation over $C_E^\varepsilon(I_1, I_2)$.

[19]Palfrey's definition can be used here as even though the parties' preferences are lexicographic they are still representable by a utility function. This is because the first two dimensions of preferences can take on only a finite number of values $(0, \frac{1}{3}, \frac{1}{2}, 1)$, and the third dimension, vote share, can be mapped into the interval [0,1]. An example of such an utility function for when the second dimension has zero weight is given by,

$$U = \left\{ \begin{array}{cc} V & \text{if } P = 0 \\ 1 + V & \text{if } P = \frac{1}{3} \\ 2 + V & \text{if } P = \frac{1}{2} \\ 3 + V & \text{if } P = 1 \end{array} \right\}$$

Similar utility functions, but with more cases, exist for when the second dimension has non-zero weight.

[20]This new definition of equilibrium is introduced in order to permit extension to the case of simultaneous competition for multiple districts. This extension is pursued later in this chapter and was explored in Chapter 1 for the case of plurality rule.

Definition 2.1 *A pair of locations, $\{I_1, I_2\}$, is a strict limit equilibrium if,*

(a) for every $y \neq I_1$, there is an $\varepsilon(y)$, such that for all $E' \in C_E^{\varepsilon(y)}(y, I_2)$ and $\tilde{E} \in C_E^{\varepsilon(y)}(I_1, I_2)$, $O_{I_1}(I_1, I_2, \tilde{E}) \succ O_{I_1}(y, I_2, E')$. And,

(b) for every $w \neq I_2$, there is an $\varepsilon(w)$, such that for all $E' \in C_E^{\varepsilon(w)}(I_1, w)$ and $\tilde{E} \in C_E^{\varepsilon(y)}(I_1, I_2)$, $O_{I_2}(I_1, I_2, \tilde{E}) \succ O_{I_2}(I_2, w, E')$.

It is possible that if $X'(I_1, I_2)$ is not a singleton then an entrant will randomize over one or several intervals. For any policy position, $x \in \Re$ and $\varepsilon(x) > 0$, define the following intervals: $x^+ = (x, x + \varepsilon(x))$, $x^- = (x - \varepsilon(x), x)$, and $x^{-+} = (x - \varepsilon(x), x + \varepsilon(x))$. If $I_1 \leq I_2$ then the entrant will locate in some subset of the intervals I_1^-, I_2^+, and z^{-+}, where $z \in (I_1, I_2)$.

As the parties have no ideological motivation in the selection of their platforms it is obvious that any equilibrium found will point to another equilibrium in which the two incumbent parties simply switch positions. Any pair of such equilibria will be considered to be the same, and so constitute just one equilibrium.

All proofs have been relegated to the appendix. Here I will present the results and an intuitive explanation. The equilibria themselves are very intuitive, the complication is in proving that no other equilibria exist.

2.3 Results

Before I present the equilibria I will provide some intuition and a preliminary result that hold for both assumptions on entrant behavior. The first intuition is that if the incumbent parties are located asymmetrically then the incumbent furthest from the median voter is in a disadvantageous position. This is easy to see if the entrant stays out of the election (because it has a zero probability of victory) as then, by the median voter theorem, the incumbent closest to the median will win the election. Significantly, and in contrast to competition under the plurality rule, this disadvantage exists even when the entrant enters. This is the case as the entrant will always locate at a point closer to the median voter than the widest incumbent, possibly even squeezing this incumbent out in the first round. Thus, this incumbent will either be eliminated in the first round or lose the run-off regardless of its opponent at that stage. This intuition is captured by the following lemma (where $P(W = j)$ is the probability that party j wins the election).

Lemma 2.1 *Suppose $|I_1| < |I_2|$. Then if E is (almost) maximizing $P(W = I_2) = 0$. That is, the incumbent furthest from the center never wins.*

To understand the logic behind the entrant's decision it is useful to consider the choices it faces. Assuming the entrant enters, if it does not squeeze I_2 out in the first round then it must be either squeezing out I_1 or maximizing its vote share. To fulfill either of these objectives the entrant will need to locate either arbitrarily close to I_1 or somewhere in between the two incumbents (as locating on the flank to the right of I_2 is dominated by locating on the flank to the left of I_1 because I_1 is closer to the median voter). Thus, even if it manages to survive the first round, I_2 will lose the run-off with certainty.

Intuitively, this lemma implies that an incumbent party that is further from the median voter has an incentive to deviate inwards. These are the same centripetal incentives faced by parties when competition is under the plurality rule.[21] However, if the incumbents are located symmetrically, and not too far apart, then the similarity with the plurality rule breaks down and the incumbent parties can no longer deviate inwards and still preclude the entrant from winning. Thus, if the incumbent parties are located asymmetrically because one party deviated towards the median from entry precluding symmetric locations, then not only can the incumbent furthest from the median not win as implied by Lemma 2.1, but neither can the incumbent closer to the median.[22]

[21] See Palfrey (1984), Callander (1999) and Chapter 1.

[22] Lemma 2.1 is not stated in this stronger form as for some distributions there may exist asymmetric incumbent platforms that preclude entry (in which case the incumbent closer to the median voter does win). However, these will not constitute equilibria as the widest party will always lose and have an incentive to deviate towards the center (see the proof of Lemma 2.3). These points require one incumbent to be relatively far from the center. Consequently these points cannot be reached by a single profitable deviation if both incumbents are close enough to the center. Therefore, if the incumbents choose symmetric positions close enough to the center they will be in equilibrium. The limit of this dispersion will be seen in the characterization of the equilibria in the next section.

To see this, suppose that the gap between the two incumbents isn't too large. The entrant will then locate just outside the incumbent closest to the median, thereby trapping this party in the middle and eliminating it in the first round. By choosing close enough to this incumbent the entrant will then be closer to the median than the other incumbent and thus win the second stage run-off. On the other hand, if the incumbents are too far apart then the entrant can locate just inside the incumbent party closest to the median, squeeze it on the outside, and then win the run-off with the other incumbent.

In contrast, if the incumbents are located symmetrically, and not too far apart, then to squeeze one incumbent in the middle the entrant must locate further from the median voter than the other incumbent, thus ensuring it loses the run-off. Combining these intuitions implies, therefore, that to prevent the entrant winning the incumbents will locate symmetrically and not too far from the median voter. Further, the incumbents do not have an incentive to deviate as this will result in an asymmetric location pair that will incite the entrant to enter and win.

The intuitions outlined here and the lemma presented above will now be employed in the following sections to characterize the equilibria of the model under the two assumptions of entrant behavior. I show that under both assumptions there exist equilibria of the model in which the incumbent parties choose symmetric policy locations. In fact, I show that all equilibria to the game must exhibit this feature.

2.3.1 Entrant Always Enters

Proposition 2.1 *If the Entrant always enters then the complete set of equilibria is characterized by the following:*

(a) There exists an infinite number of equilibria of the form $\{I_1, I_2\} = \{x, -x\}$ $\forall x \in [W^*, 0]$, *where* W^* *solves* $1 - 2F(\frac{W^*}{2}) = F(W^*)$, *and* $E \in \{I_1^-, I_2^+, \}$.[23] *For all equilibria,* $P(W = I_1) = P(W = I_2) = \frac{1}{2}, P(W = E) = 0$.

(b) Depending on F, *there may exist additional equilibria that satisfy the following necessary, but not sufficient conditions,* $\{I_1, I_2\} = \{y, -y\}$ *where* $y \in [F^{-1}(\frac{1}{4}), W^*)$ *and* $F(\frac{y}{2}) > \frac{1}{3}$. $E \in 0^{-+}$ *for all such equilibria.* $P(W = I_1) = P(W = I_2) = \frac{1}{2}, P(W = E) = 0$.

Entry affects each incumbent equally ex-ante, and so each has an equal chance of winning the election. The entrant has zero probability of winning the election. The entrant enters randomly on either of the flanks. The entrant squeezes one of the incumbents out in the first round but is then defeated by the other incumbent in the run-off. For the case of the alternative vote, where the secondary concern of parties is vote share rather than making the second round, the entrant may be eliminated in the first round if $x = W^*$. In this situation, the entrant may locate around zero as well as on the flank in order to maximize vote share. In all cases deviation by either incumbent provides scope for the entrant to win, so neither moves and there are many equilibria. Notice that if $x < F^{-1}(\frac{1}{4})$ then this violates the 'too far apart' intuition and the entrant could choose $E = x + \beta$ ($\beta > 0$) and crowd I_1 on the flank, and then defeat I_2 in the run-off as it is closer to the median.

The second group of equilibria presented above is difficult to characterize as it depends critically on the particular distribution of voters in the electorate. The first group of equilibria is independent of the particular F, as long as F satisfies the assumptions of the model. Thus, under the run-off rule there are, at the least, a

[23] *If* $x = W^*$ *and the second dimension of preferences has zero weight then* $E \in \{I_1^-, I_2^+, 0^{-+}\}$ *for all* F. *Further, if* F *is uniform then* $0^{-+} \equiv (I_1, I_2)$.

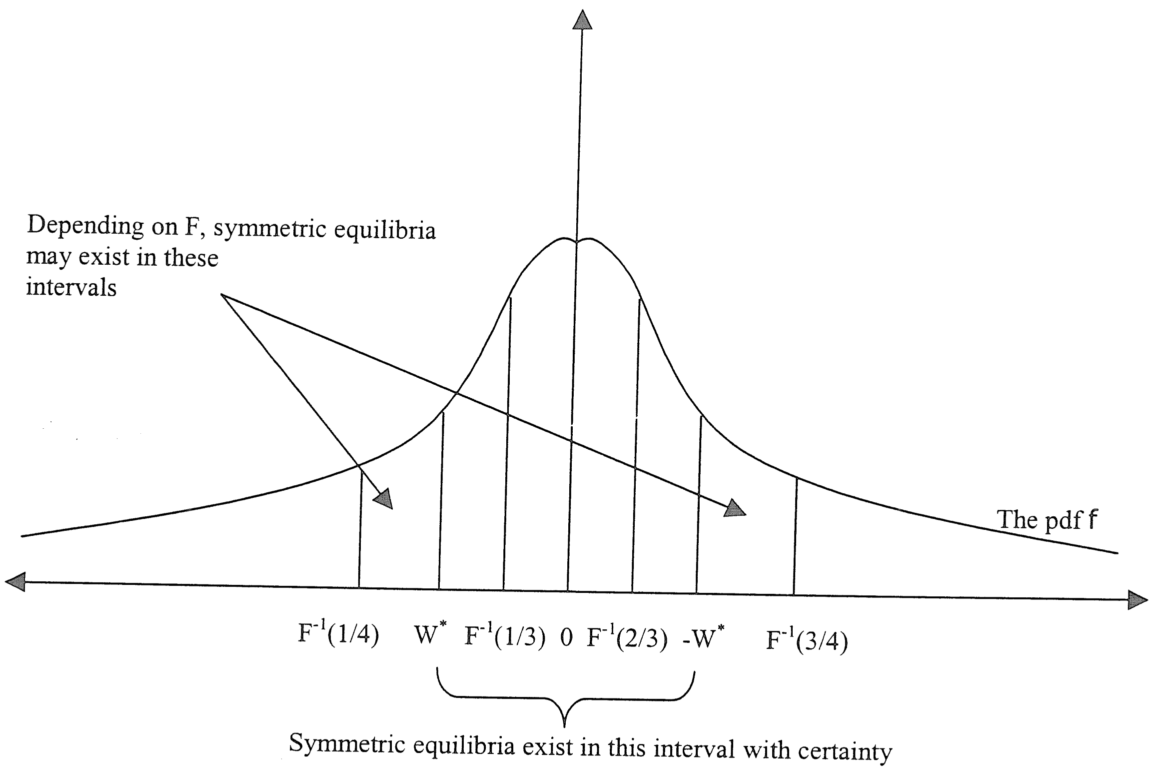


Figure 2.1: Equilibria Under the Run-Off Rule

continuum of equilibria in which the entrant never wins the election. And, independent of F, all equilibria require the incumbents to be located symmetrically about the median, and on all but a set of measure zero involve non-centrist platforms.

The party platforms described in (b) may not be equilibria as even though in the limit $E \in I_1^+$ may not lead to E having a positive probability of victory, there could still exist a point $E = I_1 + \lambda$, $\lambda > 0$, such that E has a strictly positive chance of winning the election. Whether such a point exists will depend on F and the locations of the incumbents. The incumbent locations in (a) prevent the existence of such points, which can be seen from the definition of W^*. If a pair of incumbent locations are not in the domain of (a) or (b) then such a point must exist and this location pair cannot constitute an equilibrium.

2.3.2 Enter Only if Have a Positive Probability of Victory

Proposition 2.2 *If the Entrant enters only when it has a positive probability of victory then the complete set of equilibria is characterized by the following:*

(a) There exists an infinite number of equilibria of the form $\{I_1, I_2\} = \{x, -x\}$ $\forall x \in [W^*, 0]$, *where* W^* *solves* $1 - 2F(\frac{W^*}{2}) = F(W^*)$, *and* $E = \emptyset$ *(doesn't enter). For all equilibria,* $P(W = I_1) = P(W = I_2) = \frac{1}{2}, P(W = E) = 0$.

(b) Depending on F, *there may exist additional equilibria that satisfy the following necessary, but not sufficient conditions,* $\{I_1, I_2\} = \{y, -y\}$ *where* $y \in [F^{-1}(\frac{1}{4}), W^*)$ *and* $F(\frac{y}{2}) > \frac{1}{3}$. $E = \emptyset$ *(doesn't enter).* $P(W = I_1) = P(W = I_2) = \frac{1}{2}, P(W = E) = 0$.

In all of these equilibria the entrant stays out and only two parties compete. If an incumbent deviates then, as above, the entrant could win the election. Therefore, a deviation by an incumbent would be followed by entry. This implies that the incumbents encounter the same incentives as when facing compulsory entry and, consequently, in equilibrium the incumbents choose the same positions.

2.4 Discussion

2.4.1 The Model

It is hard to say which assumption of entrant behavior is the more appropriate. There are different circumstances when each assumption would be preferred. As parties are formed primarily to win elections it may be expected that formation would only occur when there exists a chance for victory. The assumption 'enter only if have a positive probability of victory' captures these motivations.

However, in a repeated model we can certainly imagine an entrant who enters despite having zero probability of victory in the current period. Such an entrant may have aspirations for future electoral success and wish to start building a support base at the expense of other parties, or they may simply want to have a voice and feel that the cost of entry is outweighed by the value of the audience that electoral participation brings.[24] Interestingly, the location choices of the two incumbent parties are unaffected by whichever of these assumptions on entry is made.

An interesting aspect of the results found here, in relation to other electoral rules when parties are Downsian, is the large number of equilibria that exist. When this model is applied to the plurality rule there exists a unique equilibrium for 'always enter' and no pure strategy equilibrium when entrants only enter if they can win (see Palfrey (1984) and Callander (1999)). Strangely enough, this is because it is so difficult under the run-off rule for the incumbents to deter successful entry by a third party. For almost all distributions there exists, for a given location I_1, at most one location for I_2 such that successful entry is deterred (at $-I_1$). Therefore, there are no deviations available to either incumbent to make themselves better off as entry will be induced by the deviation and the entrant will win the election. This implies that

[24]The assumption that parties immediately receive the support of all voters for whom they are the closest party implies a more long term view is captured by this one shot model as difficulties of party establishment, such as name recognition, are assumed away. That is, assuming voters always vote sincerely implies that if an entrant can't win in the one shot model it won't be able to win regardless of how many periods the electoral competition is modelled as (unless, of course, its entry incites additional entry). If a dynamic model is to differ from and extend what is presented here then additional party competition for characteristics such as name recognition or platform credibility would need to be considered.

if there exists an entry deterring location I_2 for a given I_1 then this location pair is an equilibrium. Because such a point exists for any location of I_1 that isn't too far from the median, there exists a continuum of equilibria.

In contrast, under the plurality rule there exists either an empty or non-singleton set of points available to I_2 such that entry is deterred (these points are in the neighborhood of $-I_1$). Consequently, there always exist profitable deviations and, therefore, no equilibrium exists when an entrant only enters if it can win. A unique equilibrium is found under the assumption 'always enter' despite the existence of deviations that preclude successful entry. This is because deviations from the equilibrium cause the entrant, who must enter, to locate near the deviating incumbent in order to maximize its vote share, thus punishing the deviator and handing the victory to the other incumbent.

Essentially, the difference between these two rules is the ability of an entrant under the run-off rule to attack both incumbents by locating on a flank, an ability the entrant doesn't have under the plurality rule. If the incumbents are close enough together then, under both rules, entry between them results in a very low vote share and certain defeat for the entrant. Consequently, the entrant must focus on entering on the flank. Under the plurality rule such entry only affects the vote share of one incumbent. As such, the other incumbent remains unpunished and wins the election. In contrast, however, under the run-off rule entry on the flank enables the entrant to attack both incumbents. In the first round the incumbent whose flank is attacked is squeezed out. With this incumbent out of the way, the entrant is able to gain a greater vote share and attack the remaining incumbent in the second round. This second round opportunity makes it more difficult for the incumbents to deter entry and, ironically, produces far more equilibrium party locations in which successful entry is deterred.

Though the dynamic structure of the current model is rather particular, it is actually more general than it seems. One candidate can never prevent the entry of a second candidate and, therefore, if the timing of decisions is sequential or endogenous the first candidate will act as if there is a second candidate, and choose the same platform as it would if the parties instead chose simultaneously.[25,26] Also, if the potential entrant chooses to stay out then all potential entrants (who similarly consider their entry decision in isolation) would stay out. Therefore, the location pairs for the incumbents in the equilibria found here are also equilibrium location choices for the incumbent parties in a model involving an arbitrary number of potential entrants.

The established theoretical prediction for the run-off rule is what Riker (1982) refers to as "Duverger's Hypothesis" (Duverger (1954)). This assertion covers the class of electoral rules that Duverger expected to favor multi-partyism. This class incorporates the run-off and proportional representation rules as it was believed that they do not encourage parties to maximize their vote count and, therefore, the incentive to rationalize into only two parties is absent. This prediction would seem to be contradicted by the two-party equilibria found in this model. However, significantly, this is not the case given the differences in the explicit timing of the model presented here and the implicit sequencing in the reasoning of Duverger.

The two-party equilibria found here suggests that if there are currently two competing parties then a third candidate will not enter the election as it has no chance of victory (assuming the first two parties have chosen equilibrium locations). An example of this timing is the case of Australia. In Australia the alternative vote rule was established soon after the founding of the federal government and replaced the plurality rule. Therefore, parties initially made formation and entry decisions under the plurality rule and this helped establish a firm two-party system.[27] Thus, at the time the run-off rule was introduced there existed an established two-party system and there was no incentive, as explained in the current model, for a third party to enter.

[25] An endogenous timing scheme allows all parties to choose when they wish to enter in addition to the policy platform they adopt. If there exists a first point at which parties can act (the interval of time is closed) then all of the equilibria found in Section 2.3 will continue to exist. However, if such a point doesn't exist, or if timing is sequential, then the equilibrium refinement of subgame perfection will eliminate all equilibria that do not maximize the incumbent parties' vote shares.

[26] The assumption that indifferent voters would randomize over the first two parties and strictly prefer an incumbent to an entrant would need to be retained.

[27] The tendency for two-party systems to evolve under plurality rule is well documented and is known as Duverger's Law.

In contrast, the intuition of Duverger implies that if more than two parties are already competing in the electoral system then from this point they have no incentive to merge into two parties. This strategic situation is not captured by the present model. Thus the multi-partyism prediction of Duverger's Hypothesis is not at odds with the entry deterring capabilities unearthed here. The natural question, however, is why so many parties originally entered the electoral contests observed by Duverger. One possible way that this situation could arise is if the electoral rule was changed to one of run-off from something more conducive to multiple parties. Such an evolution has been the case in the French Fifth Republic.[28]

For most national elections between 1945 and 1988 France employed a dual-ballot run-off rule for both parliamentary seats and the president.[29] Significantly, however, the first system employed after the war was one of proportional representation, which induced the entry of many parties.[30] Thus, at the time the run-off rule was introduced multiple parties already existed. And as these parties did not necessarily have the incentive to then exit the electoral system France was (and still is) characterized by a multi-party system. Interestingly, though, the effective number of parties competing and succeeding in France has been diminishing through time (Lijphart (1994, p. 161)).

These possibilities imply that the predictions for the run-off rule significantly depend on the starting conditions. Therefore, a more appropriate statement of Duverger's Hypothesis may read,

> "Competition between two parties under the run-off rule is a stable configuration, though systems with more than two parties do not necessarily converge to this configuration."

Interestingly, this weaker statement about multi-party systems is consistent with

[28]See Lijphart (1994) for details. A more detailed analysis is provided by Schlesinger and Schlesinger (1990).

[29]For the presidential election only the top two candidates in the first round were permitted to compete in the run-off stage. For the parliamentary elections any candidate with more than 12.5% of the vote (the threshold was 10% for some elections) could proceed. This lower threshold would, most likely, give even greater incentive to rationalize into two parties as in this case the rule more closely resembles a plurality rule.

[30]See Cox (1997) for models and evidence of the tendency of multiple parties to enter proportional representation elections.

one of the first attempts to describe the relationship between the number of parties and the run-off rule. Lowell (1896, p. 110), in contrast to Duverger, did not claim that the use of the run-off rule is a sufficient condition for multi-partyism, even in a probabilistic sense. Instead, he only made the weaker claim that "where a number of groups exist, [the two stage majority system] tends to foster them, and prevent their fusing into larger bodies." (quoted from Wright and Riker (1989))

This conclusion has important implications for the question of electoral design. It suggests that the outcome from the introduction of an electoral rule not only depends on the characteristics of this rule, but also on the characteristics of all previous rules that have been in force in a given system. Consequently, the application of electoral theory to real world situations may be far more delicate that our simplified models would indicate.

2.4.2 Related Literature

As mentioned in the introduction, the paper closest to the current work is the model of citizen-candidates developed by Osborne and Slivinski (1996). In contrast to the current work, Osborne and Slivinksi assume candidates are policy orientated and thus, most importantly, policy restricted. They consider the opposite extreme to that modelled here by assuming candidates must choose their true ideal point as their campaign platform. They also allow the cost incurred by a party to compete in the election to be non-zero and assume the voters treat all parties identically (i.e., their analogous statement of Assumption 2.5 does not differentiate between the incumbent parties and the entrant). Osborne and Slivinski then characterize the parameters under which two-party outcomes could arise for both the run-off and plurality rules. Essentially, they find that for both rules two-party equilibria are possible if the cost of entry (relative to the benefit of winning) isn't too low to entice further entry, and isn't too high to not justify entry by two parties. They then show that two-party outcomes are more likely under the plurality rule than the run-off rule.[31]

[31]Remarkably, Osborne and Slivinski (1996) are also able to characterize conditions under which three and four party equilibria may arise, as well as the equilibria themselves.

Even with different domains of existence, the two-party equilibria found by Osborne and Slivinski (1996) under both rules have similar intuition: third party entry is deterred because the two incumbent parties are located symmetrically, and the incumbent parties can't deviate towards the median as they are policy constrained. The model presented here relaxes this final restriction and, for simplicity, considers the opposite extreme by assuming complete policy flexibility. In order to focus on the incentives that arise from this additional freedom I consider only two possible ranges of cost of entry (zero, and when two but not three parties are willing to enter if they have equal probabilities of victory) and impose the additional requirement of Assumption 2.5. This new framework alters completely the incentives faced by parties under both rules. Under the plurality rule this change has a profound impact on the set of equilibria and results in, for the two entry assumptions respectively, uniqueness and non-existences of equilibria.[32] However, remarkably, under the run-off rule these different incentives still produce a continuum of two-party non-centrist equilibria.

Another related model is that of Greenberg and Shepsle (1987).[33] Though it is sometimes construed as representing a model of run-off (see Wright and Riker (1989)), this connection is somewhat tenuous. Greenberg and Shepsle consider elections for K member districts and argue, quite appropriately, that the objective for parties in this environment is to maximize their rank rather than vote share (i.e., ensure they are in the top K vote getters). Wright and Riker (1989) argue that the case for $K = 2$ captures the incentives of a run-off election as parties want to be amongst the leading two parties to make the run-off stage. However, this argument ignores the presence of the second stage and the fact that ultimately there can be only one winner. For example, it would not seem reasonable to suppose that a party would choose a policy platform that ensured its passage to the second round if it also ensured certain defeat in this final stage. Essentially, this specification of Greenberg and Shepsle is the current model with the first dimension of preferences set to zero. However, it would seem that such a model can only capture the incentives induced by the run-off rule if

[32] See Palfrey (1984) and Callander (1999).

[33] This model has been further developed by Shvetsova (1995).

the rather unreasonable assumption is made that policy platforms chosen in the first round are ignored by voters at the run-off stage.[34]

Finally, Myerson (1993) also considers the run-off rule (in the form of the alternative vote), though his model differs significantly from the current work. He considers a distributive model of elections and doesn't allow for the prospect of entry. Myerson, therefore, does not concern himself with the equilibrium number of parties that is of interest here. Instead, he concerns himself with the equality of campaign offers that are made by parties under different voting rules.

2.5 An Extension: Multiple Districts

Often elections require political parties to compete for more than a single district. In fact, competition may be for many districts as an entire legislature is elected simultaneously. This is the case for Australian state and federal elections under the alternative vote rule. It would seem natural, therefore, to incorporate such simultaneity into models of electoral competition. In this section I extend the model to allow for multiple districts and show that the results of the single district model are robust to simultaneous competition and district heterogeneity.[35]

Immediately it can be seen that if the districts are identical then the single district equilibria hold in every district and the incumbents win all of the districts. However, if there is some heterogeneity across these districts then this may no longer be true and it becomes much more difficult for the incumbent parties to secure victory in all districts simultaneously. I find that the ability of the incumbent parties to win all districts depends critically on the degree of heterogeneity across districts and the parties' flexibility in presenting different candidate platforms in the different districts. Basically, as long as the heterogeneity of districts does not exceed the ability to differ-

[34]That is, parties are free to move their policy platforms between rounds and are not punished by voters in any way for their previous announcements. Such an assumption, of course, draws into question to a substantially greater degree the rationale behind sincere voting.

[35]This extension for the plurality rule was studied in Chapter 1. Austen-Smith (1981) and Hinich and Ordeshook (1974) also consider models of multiple districts under plurality rule, but do not incorporate the possibility of entry.

entiate individual candidates from the party platform, then the equilibria of the single district case carry over and the incumbent parties are able to deter successful entry and secure all districts for themselves. If, however, the heterogeneity is too extreme then successful third party entry cannot be prevented. This domain of successful entry deterrence can be seen as another possible explanation for why the run-off rule has produced different numbers of parties in different real world environments.

To incorporate the extension to multiple districts into the model I will make the following additional assumption.

Assumption 2.6 *There exists a continuum of districts. In district i the median voter's ideal point is Z_i. The ideal points of district median voters are distributed symmetrically about 0 on the support $[\underline{Z}, \overline{Z}]$, where $\underline{Z} = -\overline{Z}$. The distribution of voters' ideal points in district i is given by the cdf $F(x - Z_i)$ for all $x \in \Re$.*

The assumption of a continuum of districts is, of course, not realistic. However, it has been employed as it captures the effect and intuition of the multiple district scenario whilst avoiding the complexity of calculation associated with a lumpy distribution of median voters. It is in the same spirit as the assumption of a continuum of voters in the single district case.

In multiple district elections there exists both individual candidates and the larger party. Though the party itself doesn't run in individual districts (it fields representatives instead), it still stakes out a policy platform to which its individual candidates are associated. The individual candidates may be perceived somewhat differently from the party platform, but this difference is not extreme. By being a party's representative an individual candidate is associated with the party platform by voters. This association of party and candidate restricts the set of policy platforms that can be presented by an individual candidate and serves to differentiate the multiple district model from the repeated application of the single district results.

The following assumption incorporates this policy inflexibility into the model. The degree of candidate flexibility may be imposed by the party hierarchy or reflect the capabilities and talents of the individual candidates themselves. In this model, it does not matter who controls the freedom of individual candidates as the objectives of both candidate and party are aligned (and thus I will refer interchangeably to either candidate or party flexibility). However, this alignment may not always be the case, particularly under other voting rules, and an investigation of this relationship may prove fruitful in understanding the origin and nature of political parties. Let $I_j(i)$ denote the position of party j's candidate in district i (where the median voter is at Z_i).

Assumption 2.7 *I_1 and I_2 must each choose a single platform in $\Re$. The individual candidates for each party must locate within δ of these party platforms. That is, for $j = 1, 2$, $|I_j(i) - I_j| \leq \delta$ $\forall i$ such that $Z_i \in [\underline{Z}, \overline{Z}]$.*

Obviously, the size of δ indicates the degree of freedom individual candidates have in choosing their platforms. If $\delta = 0$ then candidates have no freedom and party discipline is perfect. On the other hand, if $\delta = \infty$ then there is no party discipline and candidates have complete freedom. In this case the single district results are applicable to each district separately. To differentiate the multiple district approach I will concentrate on the strategic environment when $\delta < \infty$.[36]

The order of play in this expanded game is analogous to the single district case and is as follows. The two incumbent parties choose their party and individual candidate platforms simultaneously and compete in every district. In each district a potential entrant then makes an entry decision, and if it chooses to enter it selects a platform position. The results to follow, as did the results of Section 2.3, do not depend on which entry assumption is made (always enter versus only if the entrant has a positive probability of victory). As such, throughout this section I will intentionally leave open the possibility for both types of entrant behavior. The assumption of a single entrant in each district is made for analytical simplicity and aims to capture the idea that if a single party could enter and win a district then the incumbents lose that district.

In order to understand party behavior in this expanded environment, the expanded

[36]The plurality rule results of Chapter 1 do not permit flexibility and only consider the case where $\delta = 0$. This is done, however, only for analytical simplicity and the same substantive conclusions result if the parties and candidates are permitted some flexibility.

objectives of the parties must be specified. I will assume that parties primarily seek to maximize their share of districts won. Denote this share by M_j for party j. This objective is a natural extension of the single district preferences as it requires the parties to maximize the sum of their probabilities of victory in each district.[37] The minor objectives will be similarly extended: the second preference will be to maximize the share of districts in which they make the second round (denoted by MQ_j for party j), and the third preference will be to maximize total vote share (denoted by MV_j for party j). Thus, the outcome function for party $j \in \{I_1, I_2\}$ can be written as follows.

$$O_j(I_1, I_2, E) = (M(j|I_1, I_2, E), MQ(j|I_1, I_2, E), MV_j\,(I_1, I_2, E))$$

Where $E = (..., E_i, ...)$ is the vector of locations for all entrants E_i. As each entrant competes in only one district, their preferences will remain the same as in the single district case.

The equilibrium concept provided in Definition 2.1 and employed in the single district results can also be extended naturally to the multiple district environment. In fact, the definition applies to the expanded environment with only two minor adjustments (and as such will not be restated here). Firstly, the location pair, $\{I_1, I_2\}$, is no longer a pair of points on the real line. Instead, each I_j is a vector of points specifying the location of the party platform as well as the platforms for candidates in individual districts (subject to the restriction of Assumption 2.7). Secondly, as there are many potential entrants, the choices by the entrants are given by a vector, $C^{\varepsilon}_{E}\,(I_1, I_2) = \left(..., C^{\varepsilon}_{E_i}\,(I_1, I_2)\,, ...\right)$, where $C^{\varepsilon}_{E_i}$ is the set of almost maximizing points for some entrant E_i (as defined in Section 2.2). If the pair $\{I_1, I_2\}$ satisfies the requirements of this expanded definition then I will say that it constitutes an equilibrium of the multiple district game.

The existence of multiple districts alters the strategic environment facing the

[37]There are many alternative specifications of party objective functions that could be employed. Perhaps the most plausible would be to maximize the probability of winning government. Unfortunately, as governments can be formed with a minority of seats, or by forming a coalition of parties, a complex model of government formation would need to be incorporated for this assumption to be used.

incumbent parties. With their choice of policy platform in each district restricted, it will now be more difficult for them to prevent successful third party entry. The following lemma characterizes exactly the degree of this capability in the presence of district heterogeneity.

Lemma 2.2 *Suppose the incumbent party platforms are at* I_1 *and* I_2*, and that* $I_1(i) = I_1$ *and* $I_2(i) = I_2$ *in district* i *where* $Z_i = \left(\frac{I_1+I_2}{2}\right)$*. Then if successful entry can be deterred in district* i *it can also be deterred in all districts with median voters in the interval* $\left[\left(\frac{I_1+I_2}{2}\right) - \delta, \left(\frac{I_1+I_2}{2}\right) + \delta\right]$.

To understand this lemma consider the situation in which the incumbent parties are located symmetrically around the median voter of the central district. If the incumbents are not too far from the median then they will deter successful third party entry by locating their candidates at the same points as their party platforms (or, indeed, any pair of symmetric locations not to far from the median).[38] Lemma 2.2 implies that they can also deter entry in districts with medians in the interval $[-\delta, \delta]$. This is possible as in such districts the two incumbent parties can field candidates away from their party platforms such that they are symmetric in each district, thus deterring successful entry. However, for districts with median voters outside of this domain the incumbent parties cannot achieve symmetry and successful entry may be possible. Obviously this logic is not contingent upon the incumbents locating symmetrically about the median voter of the central district, and the lemma holds for some interval, regardless of the location of the incumbents.

Lemma 2.2 characterized the entry deterring capability of the incumbent parties. The following lemma strengthens this result and shows that requiring the interval of district medians to be a subset of this interval is not only sufficient for entry deterrence, but also necessary if the incumbent location pair is to constitute an equilibrium.

Lemma 2.3 *If* $|\overline{Z}| = |\underline{Z}| > \delta$ *then there does not exist an equilibrium in which successful third party entry is deterred in all districts.*

[38] The bounds of Propositions 2.1 and 2.2 characterize the bounds for entry deterrence in any given district.

This lemma shows that if the heterogeneity of districts is too great relative to the flexibility of the party platforms then in equilibrium successful third party entry is inevitable.[39] If, on the other hand, this condition is not satisfied then Lemma 2.2 implies that successful entry can be deterred in every district if incumbent party platforms are located appropriately. The following proposition shows in this case that not only can entry be deterred, but that some of these location pairs constitute equilibria. Define I_1^* to be the minimum I_1 such that $\{I_1, I_2\} = \{I_1^*, -I_1^*\}$ is an equilibrium to the single district game that deters successful entry.

Proposition 2.3 *Suppose that the pair of party platforms,* $\left\{\tilde{I}_1, \tilde{I}_2\right\}$ *is an equilibrium to the single district game in which successful entry is deterred, and that* $\left|\overline{Z}\right| = \left|\underline{Z}\right| \leq \delta$. *Then the following are equilibria to the multiple district game:*

$$\begin{aligned} \{I_1, I_2\} &= \left\{\tilde{I}_1, \tilde{I}_2\right\} \\ I_1(i) &= \tilde{I}_1 + Z_i + \varepsilon_i \\ I_2(i) &= \tilde{I}_2 + Z_i + \varepsilon_i \\ \forall \varepsilon_i &\in [\delta - Z_i, \delta + Z_i], \ \textit{subject to } \tilde{I}_1 + \varepsilon_i \leq I_1^* \end{aligned}$$

In these equilibria the two incumbent parties deter successful entry and win all districts. If entrants only enter if they have a positive probability of victory then $E_i = \emptyset$ *for all* i. *If entrants always enter then they locate in each district as specified in Proposition 2.1.*

Further, if $\left|\overline{Z}\right| = \left|\underline{Z}\right| < \delta$ *then there exists additional equilibria involving asymmetric party platforms (i.e., where* $|I_1| \neq |I_2|$*).*

This proposition states that the single district results are robust to some district heterogeneity. In fact, the amount of heterogeneity that can be tolerated depends critically on the flexibility of individual candidates to differentiate themselves from

[39] If $\delta = 0$ then a stronger result can be stated: if $\left|\overline{Z}\right| = \left|\underline{Z}\right| > 0$ then there does not exist an incumbent location pair such that successful entry is deterred in every district, regardless of whether this location pair constitutes an equilibrium to the multiple district game (see Callander (1999) for a proof of this result).

the party platform. These equilibria require the individual candidates to differentiate themselves from the party platforms such that they are symmetrically located with their opposing candidate in their respective districts. Basically, if the heterogeneity isn't too extreme then the single district equilibrium can be achieved in every district. The second part of the proposition points out that if the candidate flexibility exceeds the district heterogeneity then symmetric positioning in each district can be achieved even if the parties themselves have asymmetric platforms.[40]

The equilibria of the multiple district case build upon the intuition of the single district case. In the single district case a continuum of equilibria exist precisely because it is so difficult for the incumbents to prevent successful third party entry. The requirement that parties compete simultaneously for many districts makes this task even harder. If the incumbent parties have enough flexibility in the platforms they can offer in each district then they can overcome this difficulty and still prevent successful third party entry. However, if the district heterogeneity is too great then the task of entry deterrence reaches the point of impossibility and successful third party entry is inevitable.

These results can, like the single district results, be compared to the equilibria of the model under the plurality rule derived in Chapter 1. It was pointed out in Section 2.4 that in a single district under the plurality rule there are, in a sense, too many incumbent location pairs that deter successful entry. Consequently, profitable deviations always exist for the incumbent parties and equilibria fail to exist if the entrant can choose to stay out. As is the case with the run-off rule, the extension to multiple districts makes successful entry deterrence more difficult under the plurality rule. However, it makes it more difficult in such a way that a unique, two-party entry deterring equilibrium exists as long as heterogeneity is not too extreme (and this bound is far larger than under the run-off rule: for $\delta = 0$ the two-party equilibrium exists under the plurality rule if $|\underline{Z}| \leq Z^*$ where $f(Z^*) = \frac{1}{3}$, whereas it fails to exist under the run-off rule for the same δ if there is any district heterogeneity). Thus, in

[40]This, of course, is not a complete characterization of equilibria to the multiple district game. Though such a characterization is possible, it would not add substantively to the results and has therefore been omitted.

comparison, district heterogeneity makes successful entry deterrence generally more difficult for the incumbents. Under the plurality rule this difficulty translates to stable equilibrium outcomes. In contrast, however, under the run-off rule this difficulty can very quickly become overwhelming and break down the stable two-party equilibria that would otherwise exist.

2.6 Conclusion

In this chapter, I characterize the equilibria to a model of electoral competition with entry under the run-off rule. The primary finding is that a continuum of equilibria exist in which one of two incumbent parties always win. This result is found to be robust to variations in the motivations of the entrant, the timing of entry decisions, as well as the preferences of the parties. I also consider an extension of the model to simultaneous competition for multiple districts and find that the equilibria of the single district case are robust to some district heterogeneity, and characterize the sufficient and necessary bounds on this heterogeneity such that two-party equilibria exist.

The inability of the third party to win the election implies that only two parties will enter and compete in the election if parties are victory seeking. In contrast, "Duverger's Hypothesis" asserts that if there are multiple parties already competing then the incentive to rationalize into only two parties will be absent. Interestingly, both of these assertions may be simultaneously true. This possibility suggests that the predictions of the model are sensitive to the initial conditions of the electoral system, implying that the application of theoretical models to real world situations may be more delicate than is allowed for by current specifications.

2.7 Appendix

In the text I indicated that the preferences of the parties may vary depending on whether the run-off rule is describing competition under the dual-ballot or the alternative vote systems (the difference is whether the second dimension of preferences receives positive weight). Despite this, the proofs of all the results vary minimally if the parties have two or three dimensions of preference (that is, whether they derive utility from reaching the second round of counting or balloting). Here I will present the proofs for when parties have only two dimensions of preference. If an additional argument is needed for the three dimensional case I will mark an * and include these arguments in a footnote.

As is standard, all proofs will proceed by showing there exists profitable deviations from any candidate locations other than those claimed to constitute equilibria. I will denote the locations under consideration by I_1, I_2, and E, and any deviation with a tilde (e.g., $\tilde{I}_1$). For simplicity, if the arguments of a function are I_1, I_2, and E then they will be omitted. In an abuse of notation let $E_i = I_j^+$ denote an entrant in district i locating arbitrarily close to the right of incumbent party j. WOLOG I will assume throughout that if $I_1 \neq I_2$ then $I_1 < I_2$. Vote share refers to primary vote share.

2.7.1 Proof of Lemma 2.1

If there are only two parties remaining in the election (either the entrant didn't enter or one party has been eliminated) and these parties are λ and κ, where $\lambda, \kappa \in \{I_1, I_2, E\}$, then it is easy to see that $P(W = \lambda) = 1$ iff $|\lambda| < |\kappa|$ and $P(W = \lambda) = P(W = \kappa) = \frac{1}{2}$ iff $|\lambda| = |\kappa|$ (that is, the party closest to the median voter will win). Therefore, for $P(W = I_2) > 0$ it must be that $|E| \geq |I_2|$ (as otherwise I_2 will be the furthest from the median voter regardless of which other party makes it through the first round).

Suppose then that $E \geq I_2$. This implies that $V_{I_1} = F(\frac{I_1+I_2}{2}) > \frac{1}{2}$. Therefore I_1 survives the first round and as $|I_1| < |I_2|, |E|$ it must be that $P(W = I_1) = 1$. So instead suppose that $E \leq -I_2$. I will proceed by showing that all such locations are dominated by $\tilde{E} = I_1^-$ and, therefore, the entrant can't be (almost) maximizing at any point such that $P(W = I_2) > 0$. To show this I will consider vote shares for all parties at the different entrant locations. For the entrant, $V_E = F(\frac{E+I_1}{2})$ and $V_E(I_1, I_2, \tilde{E}) \rightarrow$

$F(I_1) > F(\frac{E+I_1}{2})$. For I_1, $V_{I_1} = F(\frac{I_1+I_2}{2}) - F(\frac{E+I_1}{2})$ and $V_{I_1}(I_1, I_2, \tilde{E}) \to F(\frac{I_1+I_2}{2}) - F(I_1) > V_{I_1}$. The vote share for I_2 is unaffected. Therefore, the entrant's vote share strictly increases, the incumbent parties' vote shares weakly decrease, and the entrant moves closer to the median voter. Consequently, $O_E(I_1, I_2, \tilde{E}) \succ O_E(I_1, I_2, E)$ and $P(W = I_2) = 0$ if the entrant is (almost) maximizing.

2.7.2 Proof of Proposition 2.1: Entrant Always Enters

Define $W' = [W^*, -W^*]$, where W^* solves $F(W^*) = 1 - 2F(\frac{W^*}{2})$. From the proof of Lemma 2.1, $E = I_1^-$ is the dominant choice for $\forall E < I_1^-$. Similarly for $E = I_2^+$. Therefore, in the proofs to follow I will only consider values of $E \in \{I_1^-, I_2^+, (I_1, I_2)\}$.

Case 1 $I_1, I_2 \in W'$.

Case 1.1: $I_1 = -I_2$.

If $E \in (I_1, I_2)$ then $V_E < 1 - 2F(\frac{W^*}{2})$ and $V_{I_1}, V_{I_2} > F(W^*)$, which imply that E loses in the first round. If $E \in \{I_1^-, I_2^+\}$ then $|E| > |I_1|, |I_2|$ and so E will not win the run-off. Thus, it must be that $P(W = E) = 0$. If $I_1 \neq W^*$ then $C_E^\varepsilon = \{I_1^-, I_2^+\}$. If $x = W^*$ and F isn't uniform then $C_E^\varepsilon = \{I_1^-, I_2^+, 0^{-+}\}$.*[41] If $I_1 = W^*$ and F is uniform then $C_E^\varepsilon = \{I_1^-, I_2^+, (I_1, I_2)\}$. In all circumstances the incumbents are affected equally and $P(W = I_1) = P(W = I_2) = \frac{1}{2}$.

Consider a deviation from symmetry by I_1 to $\tilde{I}_1$. If $\left|\tilde{I}_1\right| > |I_1|$ then by Lemma 2.1 their probability of victory is zero and $O_{I_1} \succ O_{I_1}(\tilde{I}_1, I_2, E)$. So suppose that $\tilde{I}_1 \in (I_1, I_2]$. If $E = \tilde{I}_1^-$ then $V_{I_1}(\tilde{I}_1, I_2, E) < 1 - 2F(\frac{W^*}{2})$ as $E, I_2 \in W'$. Further, as $V_E = F(\tilde{I}_1) > F(W^*)$ and $V_{I_2} = 1 - F(\frac{\tilde{I}_1+I_2}{2}) > F(W^*)$ it must be that $V_E, V_{I_2} > V_{\tilde{I}_1}$ and $P(W = \tilde{I}_1|\tilde{I}_1, I_2, E) = 0$. This implies that $P(W = E|\tilde{I}_1, I_2, E) = 1$ as $|E| < |I_2|$. Thus, for the optimal choice by the entrant it must be that $P(W = \tilde{I}_1|\tilde{I}_1, I_2, E) = 0$ and $O_{I_1} \succ O_{I_1}(\tilde{I}_1, I_2, E)$. Thus, $\{I_1, I_2\} = \{y, -y\}$ where $y \in W'$ is a strict Nash equilibrium.

Case 1.2: $|I_1| < |I_2|$.

[41] If the second dimension of preferences has positive weight then $E \in \{I_1^-, I_2^+\}$ and the same argument holds.

From Lemma 2.1 it must be that if E is almost maximizing then $P(W = I_2) = 0$. Therefore, these asymmetric locations can't constitute an equilibrium as $\tilde{I}_2 = -I_1 \implies P(W = I_1) = \frac{1}{2}$ and $O_{I_2} \prec O_{I_2}(I_1, \tilde{I}_2, E)$.

Case 2 $I_1 \in W', I_2 \notin W'$.

This implies that $|I_2| > |I_1|$, and so $P(W = I_2) = 0$. If $\tilde{I}_2 = -I_1$ then $I_1, \tilde{I}_2 \in W'$ and so, by Case 1.1, $P(W = I_2 | I_1, \tilde{I}_2, E) = \frac{1}{2}$. Thus, $O_{I_2} \succ O_{I_2}(I_1, \tilde{I}_2, E)$.

Case 3 $I_1, I_2 \notin \overline{W}$.

Firstly, define $\overline{W} = [F^{-1}(\frac{1}{4}), F^{-1}(\frac{3}{4})]$. This case provides examples of situations where an entrant does not maximize its utility my maximizing its primary vote share.

Case 3.1: $I_1, I_2 < F^{-1}(\frac{1}{4})$.

Firstly, notice that $E \in I_2^+ \implies P(W = E) = 1$.

Case 3.1.1: $I_1 = I_2 < 0$. In this case $C_E^{\varepsilon}(I_1, I_2) = I_2^+$ and $P(W = E | E \in C_E^{\varepsilon}(I_1, I_2)) = 1$. Further, $P(W = I_1) = 0$ and as $\varepsilon \to 0, V_{I_1} \to \frac{1}{2}F(I_1)$. If $F(I_1) \neq 0$ then as F is atomless, δ small enough can be found s.t. $\tilde{I}_1 = I_1 - \delta \implies V_{I_1}(\tilde{I}_1, I_2, E) = F(I_1 - \frac{\delta}{2}) > \frac{1}{2}F(I_1)$ and therefore $O_{I_1} \prec O_{I_1}(\tilde{I}_1, I_2, E)$. If $F(I_1) = 0$ then as F is non-degenerate a γ small enough can be found s.t. $0 < F(I_2 + \gamma) < \frac{1}{2}$. Given this, the deviation $\tilde{I}_2 = I_2 + \gamma \implies C_E^{\varepsilon}(I_1, I_2) = I_2^+$ and $V_{I_2}(I_1, \tilde{I}_2, E) = F(\tilde{I}_2) - F(\frac{I_1 + \tilde{I}_2}{2}) > 0$, as F is strictly increasing once $F > 0$. Consequently, $O_{I_2} \prec O_{I_2}(I_1, \tilde{I}_2, E)$.

Case 3.1.2: $I_1 \neq I_2$. As in the previous subcase, $C_E^{\varepsilon}(I_1, I_2) = I_2^+$ and $P(W = E | E \in C_E^{\varepsilon}(I_1, I_2)) = 1$. Further, $P(W = I_1) = 0$ and $V_{I_1} = F(\frac{I_1 + I_2}{2})$. If $F(I_2) \neq 0$ then as F is atomless there exists an α small enough such that $\tilde{I}_1 = I_2 - \alpha \implies V_{I_1}(\tilde{I}_1, I_2, E) > V_{I_1}$. If $F(I_2) = 0$ then consider the same deviation as in the subcase above. For both possibilities $O_{I_1} \prec O_{I_1}(\tilde{I}_1, I_2, E)$.

Case 3.2: $I_1 < F^{-1}(\frac{1}{4}), I_2 > F^{-1}(\frac{3}{4})$.

Let $|I_1| \leq |I_2|$ and consider possible locations for the entrant. If $E > I_2$ then it is the furthest from the median voter and $P(W = E) = 0$. If $E \in I_1^-$ then $V_{I_1} \to F(\frac{I_2 + I_1}{2}) - F(I_1) > \frac{1}{4} > F(I_1) \leftarrow V_E$ and thus $P(W = E) = 0$. Consider then $E = F^{-1}(\frac{1}{4})$. This implies that $V_{I_1} = F[\frac{I_1 + F^{-1}(\frac{1}{4})}{2}] < \frac{1}{4}$ and $V_E =$

$F[\frac{I_2+F^{-1}(\frac{1}{4})}{2}] - F[\frac{I_1+F^{-1}(\frac{1}{4})}{2}] > \frac{1}{4}$. Therefore, the entrant must survive the first round and as $|E| < |I_1|, |I_2|$ the entrant will win the run-off against whoever survives. Consequently, $P(W = E) = 1$. These results imply that to maximize the entrant must locate between the incumbents. I will now establish where the entrant will locate to maximize its primary vote share.*[42]

For a given E, $V_E = F(\frac{I_2+E}{2}) - F(\frac{E+I_1}{2})$. Differentiating this expression indicates that to maximize vote share the entrant will locate such that $f(\frac{I_2+E}{2}) = f(\frac{E+I_1}{2})$, which results in $V_{I_1} = V_{I_2}$. If such an maximum doesn't exist then it is easy to see that the entrant locates at I_1^+. If vote share is maximized at I_1^+ then from the expression above $V_E > V_{I_1}, V_{I_2}$. Thus $P(W = E) = 1$ and $C_E^{\varepsilon}(I_1, I_2) = I_1^+$. Similarly, if the vote share is maximized at a point x such that $V_{I_1} = V_{I_2}$ and $V_E > \frac{1}{3}$ then $C_E^{\varepsilon}(I_1, I_2) = x^{-+}$. However, if at such a point x, $V_E \leq \frac{1}{3}$ then $C_E^{\varepsilon}(I_1, I_2) \neq x$ (as $P(W = E) < 1$). To secure victory the entrant must sacrifice some vote share and move towards one of the incumbents. Essentially, the entrant attacks one of the incumbents, eliminating it in the first round, but remaining close enough to the median voter to win the run-off. From the expression for V_E, the entrant will be giving up vote share as it moves closer to an incumbent. As $E = F^{-1}(\frac{1}{4})$ ensures victory for the incumbent, by the continuity of F there must exist a point $K \in (F^{-1}(\frac{1}{4}), x)$ such that $V_E = V_{I_1}$. Therefore, $K^- \in C_E^{\varepsilon}(I_1, I_2)$. An analogous argument shows that a similar point exists of the form $\overline{K} \in (x, F^{-1}(\frac{3}{4}))$ and that $C_E^{\varepsilon}(I_1, I_2) = \left\{K^-, \overline{K}^+\right\}$. Note that for any location choice for the entrant in this situation it will be that $V_{I_1} \neq V_{I_2}$. However, as the entrant randomizes over these choices it will be that $E(V_{I_1}) = E(V_{I_2})$ (in expectation).

So far I have established how E will react to location choices of incumbents. I am now able to consider possible deviations for the incumbents and determine the outcomes they will produce. It suffices to consider only the vote share of the deviating incumbent as prior to deviation $P(W = I_1) = P(W = I_2) = 0$.*[43] I will denote the

[42]As E wins it definitely makes the second round, so the same argument holds for the different preferences.

[43]Only if $E \in I_1^+$ does I_1 make the second round for sure, otherwise the two incumbents have equal chances of making it. Therefore, any inward deviation by an incumbent such that $\tilde{I}_1, \tilde{I}_2 \notin \overline{W}$

entrant's platform choice after a deviation by an incumbent by E'.

Case 3.2.1: $|I_1| < |I_2|$. Consider the deviation $\tilde{I}_2 = -I_1$. After this deviation E will maximize vote share by locating at zero. With this deviation in mind consider the original subcases.

Case 3.2.1.a: $C_E^\varepsilon = I_1^+$ (thus, $V_{I_1} > V_{I_2}$). If $E' = 0 \Longrightarrow V_{I_1}(I_1, \tilde{I}_2, E') = V_{I_2}(I_1, \tilde{I}_2, E') < \frac{1}{3}$ then $C_E^\varepsilon(I_1, \tilde{I}_2, E') = 0^{-+}$. As $(I_1, \tilde{I}_2) \subset (I_1, I_2)$ then $V_E(I_1, \tilde{I}_2, E') < V_E(I_1, I_2, E') < V_E(I_1, I_2, E)$, and so $V_{I_2}(I_1, \tilde{I}_2) > V_{I_2}$ (because $V_{I_1}(I_1, \tilde{I}_2) + V_{I_2}(I_1, \tilde{I}_2) > V_{I_1} + V_{I_2}$, $V_{I_1} > V_{I_2}$, and $V_{I_1}(I_1, \tilde{I}_2) = V_{I_2}(I_1, \tilde{I}_2)$). Thus, $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$. If $E' = 0 \Longrightarrow V_{I_1} = V_{I_2} \geq \frac{1}{3}$ then $E \in \{K^-(I_1, \tilde{I}_2), \overline{K}^+(I_1, \tilde{I}_2)\}$. Similarly to above, as $(I_1, \tilde{I}_2) \subset (I_1, I_2)$ then $V_E(I_1, \tilde{I}_2, E') < V_E(I_1, I_2, E') < V_E(I_1, I_2, E)$, and so $E[V_{I_2}(I_1, \tilde{I}_2)] > V_{I_2}$. Thus, $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

Case 3.2.1.b: $C_E^\varepsilon = x^{-+}$ (thus, $V_{I_1} = V_{I_2}$). If $E' = 0 \Longrightarrow V_{I_1} = V_{I_2} < \frac{1}{3}$ then $C_E^\varepsilon = 0^{-+}$. As above, $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$. If $E = 0 \Longrightarrow V_{I_1} = V_{I_2} \geq \frac{1}{3}$ then $E \in \{K^-(I_1, \tilde{I}_2), \overline{K}^+(I_1, \tilde{I}_2)\}$ and, as above, $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

Case 3.2.1.c: $C_E^\varepsilon = \{K^-, \overline{K}^+\}$ (thus, $E[V_{I_1}] = E[V_{I_2}]$; i.e., E attacks one incumbent). This requires that $E' = 0 \Longrightarrow V_{I_1}(I_1, \tilde{I}_2, E') = V_{I_2}(I_1, \tilde{I}_2, E') \geq \frac{1}{3}$ and so $C_E^\varepsilon(I_1, \tilde{I}_2) = \{K^-(I_1, \tilde{I}_2), \overline{K}^+(I_1, \tilde{I}_2)\}$. As $(I_1, \tilde{I}_2) \subset (I_1, I_2)$ then $V_E(I_1, \tilde{I}_2) < V_E(I_1, I_2 | E') < V_E$. As in expectation it is still the case that $V_{I_1}(I_1, \tilde{I}_2) = V_{I_2}(I_1, \tilde{I}_2)$ then $E[V_{I_2}(I_1, \tilde{I}_2)] > E[V_{I_2}]$ and $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

Case 3.2.2: $|I_1| = |I_2|$. As $E = 0 \Rightarrow V_{I_1} = V_{I_2}$ then $C_E^\varepsilon \subseteq \{0^{-+}, K^-, \overline{K}^+\}$. I will now consider these possibilities in turn.

Case 3.2.2.a: $C_E^\varepsilon = 0^{-+}$ (thus, $V_{I_1} = V_{I_2} < \frac{1}{3}$). Therefore, there exists a γ small enough such that if $\tilde{I}_2 = I_2 - \gamma$ then $E' = (\frac{\gamma}{2})^{-+}$ implies $V_{I_1}(\tilde{I}_1, I_2) = V_{I_2}(\tilde{I}_1, I_2) < \frac{1}{3}$. Thus, $C_E^\varepsilon = (\frac{\gamma}{2})^{-+}$. As $(I_1, \tilde{I}_2) \subset (I_1, I_2)$ then $V_E(I_1, \tilde{I}_2) < V_E$. This implies that $V_{I_2}(\tilde{I}_1, I_2) > V_{I_2}$ and, therefore, $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

Case 3.2.2.b: $C_E^\varepsilon = \{K^-, \overline{K}^+\}$ (thus, $E[V_{I_1}] = E[V_{I_2}] \geq \frac{1}{3}$). Suppose that $\tilde{I}_2 = I_2 - \delta$ such that $\tilde{I}_2 \notin \overline{W}$. Then, as $(I_1, \tilde{I}_2) \subset (I_1, I_2)$, for small enough δ it must be that $E = \frac{\delta}{2} \Longrightarrow V_{I_1}(\tilde{I}_1, I_2) = V_{I_2}(\tilde{I}_1, I_2) \geq \frac{1}{3}$. Therefore, $C_E^\varepsilon(I_1, \tilde{I}_2) =$

implies that the probability of making the second round can only increase. Therefore, if the vote share of the deviating incumbent increases then it must be strictly better off.

$\{K^-(I_1,\tilde{I}_2),\overline{K}^+(I_1,\tilde{I}_2)\}$. Also, as $(I_1,\tilde{I}_2)\subset(I_1,I_2)$ it must be that $V_E(\tilde{I}_1,I_2)<V_E$ and $E[V_{I_2}(\tilde{I}_1,I_2)]>E[V_{I_2}]$ (as $E[V_{I_1}(\tilde{I}_1,I_2)]=E[V_{I_2}(\tilde{I}_1,I_2)]$). Thus, $O_{I_2}(I_1,\tilde{I}_2,E')\succ O_{I_2}$.

Case 4 $I_1,I_2\in\overline{W}/W'$.

Note that the region $\overline{W}/W'$ may be empty. For example, if f is uniform on $[-\frac{1}{2},\frac{1}{2}]$ then $\overline{W}=W'$. I will show that if $I_1\neq -I_2$ or $F(\frac{I_1}{2})<\frac{1}{3}$ when $I_1=-I_2$ then $\{I_1,I_2\}$ can't constitute an equilibrium.

Case 4.1: $I_1,I_2<W^*$.

This subcase employs the same deviations and analysis as Case 3.1.

Case 4.2: $I_1\in[F^{-1}(\frac{1}{4}),W^*)$, $I_2\in(-W^*,F^{-1}(\frac{3}{4})]$.

Case 4.2.1: $I_2=-I_1$ and $F(\frac{I_1}{2})<\frac{1}{3}$. $E=0\Longrightarrow V_{I_1}=V_{I_2}<\frac{1}{3},V_E>\frac{1}{3}$, and as $|E|<|I_1|,|I_2|$ it must be that $P(W=E)=1$. By the arguments of Case 3, $C_E^\varepsilon=0^{-+}$. Consider the deviation $\tilde{I}_1=I_1+\gamma$ such that $F(\frac{\tilde{I}_1}{2})<\frac{1}{3}$. Then the same arguments from Case 3 imply that $C_E^\varepsilon(\tilde{I}_1,I_2,E')=(-\frac{\gamma}{2})^{-+}$, $V_{I_1}(\tilde{I}_1,I_2,E')=V_{I_2}(\tilde{I}_1,I_2,E')$, and $P(W=E)=1$. However, this also implies that $V_E(\tilde{I}_1,I_2,E')<V_E$ and so $V_{I_1}(\tilde{I}_1,I_2,E')>V_{I_1}$.*[44] Thus, $O_{I_1}(\tilde{I}_1,I_2,E')\succ O_{I_1}$.

Case 4.2.2: $I_2=-I_1$ and $F(\frac{I_1}{2})=\frac{1}{3}$. By the symmetry of f, $E=0\Rightarrow V_{I_1}=V_{I_2}=V_E=\frac{1}{3}$. As f is continuous there exists an $\varepsilon>0$ small enough such that $E=\varepsilon\Rightarrow V_E>V_{I_2}$ and as $|E|<|I_1|,|I_2|$, it must be that $P(W=E)=1$. Therefore, $C_E^\varepsilon=\left\{K^-,\overline{K}^+\right\}$. For a small enough γ, $\tilde{I}_1=I_1+\gamma\Rightarrow C_E^\varepsilon(\tilde{I}_1,I_2)=\{K^-(\tilde{I}_1,I_2),\overline{K}^+(\tilde{I}_1,I_2)\}$ (as $V_E(\tilde{I}_1,I_2)$ is arbitrarily close to $\frac{1}{3}$ and $F(\tilde{I}_1)$ is strictly less than $\frac{1}{3}$). As $(\tilde{I}_1,I_2)\subset(I_1,I_2)$ then $E\left[V_{I_1}\left(I_1,I_2,E'\right)\right]>E\left[V_{I_1}\right]$ and $O_{I_1}(\tilde{I}_1,I_2,E')\succ O_{I_1}$.

Case 4.2.3: $I_2=-I_1+\alpha$, where $\alpha>0$, and $F(\frac{I_1}{2})\leq\frac{1}{3}$. Lemma 2.1 implies that in equilibrium $P(W=I_2)=0$. By the arguments of Case 3 it must be that $C_E^\varepsilon\subset(I_1,I_2)$, $P\left(W=E\right)=1$, and $V_{I_1}\geq V_{I_2}$. Likewise, if $\tilde{I}_2=-I_1$ then $C_E^\varepsilon(I_1,\tilde{I}_2,E')=0^{-+}$, $P(W=E|I_1,\tilde{I}_2,E')=1$, but $V_{I_1}(I_1,\tilde{I}_2,E')=V_{I_2}(I_1,\tilde{I}_2,E')$. Therefore, as $(I_1,\tilde{I}_2)\subset(I_1,I_2)$ it must be that $V_E(I_1,\tilde{I}_2)<V_E$.*[45] This implies that $V_{I_2}(\tilde{I}_1,I_2,E')>V_{I_2}$ and so $O_{I_2}(I_1,\tilde{I}_2,E')\succ O_{I_2}$.

[44]For both I_1 and $\tilde{I}_1$ the incumbents tie in the first round so the probability of making the second round is unaffected by the deviation.

[45]For $\tilde{I}_2$ the probability of making the second round is now $\frac{1}{2}$, which is at least as great as when at I_2.

Case 4.2.4: $I_2 = -I_1 + \alpha$, where $\alpha > 0$, and $F(\frac{I_1}{2}) > \frac{1}{3}$. Lemma 2.1 implies that $P(W = I_2) = 0$. Consider the deviation $\tilde{I}_2 = -I_1$. If $C_E^{\varepsilon}(I_1, \tilde{I}_2, E')$ is such that $P(W = I_2|I_1, \tilde{I}_2, E') > 0$ then $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$ and $\{I_1, I_2\}$ can't constitute an equilibrium. So suppose then that $P(W = I_2|I_1, \tilde{I}_2, E') = 0$. As $E' \notin (I_1, \tilde{I}_2)$ implies that $|E'| > |I_1|, \left|\tilde{I}_2\right|$, and $E' = 0 \Rightarrow V_{I_1}(I_1, \tilde{I}_2, E') = V_{I_2}(I_1, \tilde{I}_2, E') > \frac{1}{3}$, then it must be that $C_E^{\varepsilon}(I_1, \tilde{I}_2) = \{K^-(I_1,\tilde{I}_2), \overline{K}^+(I_1,\tilde{I}_2)\}$, where K^- and $\overline{K}^+$ are as defined in Case 3 (as otherwise the two incumbents win with probability $\frac{1}{2}$ each). Consider the entrant location $E' \in K^-(I_1,\tilde{I}_2)$. As $(I_1, \tilde{I}_2) \subset (I_1, I_2)$ then it must be that $V_E(I_1,\tilde{I}_2, E') < V_E(I_1,I_2, E')$. This implies that $I_1^- \notin C_E^{\varepsilon}$ because if such a location is not optimal after the deviation of I_2 then it couldn't have been optimal before. Therefore, before the deviation the entrant must have proved victorious by attacking one of the incumbents or locating arbitrarily close to I_1 (see Case 3). Combining the previous two observations implies that $E[V_{I_1}(I_1,I_2, E')] > E[V_{I_1}]$ and $E[V_{I_1}] \geq E[V_{I_2}]$. Finally, because $C_E^{\varepsilon}(I_1, \tilde{I}_2) = \{K^-(I_1,\tilde{I}_2), \overline{K}^+(I_1,\tilde{I}_2)\} \Rightarrow E[V_{I_1}(I_1, \tilde{I}_2, E')] = E[V_{I_2}(I_1, \tilde{I}_2, E')]$, then it must be that $E[V_{I_2}(I_1, \tilde{I}_2, E')] > E[V_{I_2}]$ and $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.*[46]

Case 5 $I_1 \notin \overline{W}, I_2 \in \overline{W}/W'$.

<u>*Case 5.1:*</u> $I_1, I_2 < 0$. Therefore, $C_E^{\varepsilon} = I_2^+$ which implies that $V_E > \frac{1}{2}$ and $P(W = E) = 1$. As f is atomless there exists a τ small enough such that $\tilde{I}_1 = I_2 - \tau$ implies $C_E^{\varepsilon}(\tilde{I}_1, I_2, E') = I_2^+$ and $V_{I_1}(\tilde{I}_1, I_2, E') > V_{I_1}$.*[47] Thus, $O_{I_1}(\tilde{I}_1, I_2, E') \succ O_{I_1}$.

<u>*Case 5.2:*</u> $I_1 < F^{-1}(\frac{1}{4})$, and $I_2 \in (W^*, F^{-1}(\frac{3}{4})]$. Lemma 2.1 implies that $P(W = I_2) = 0$. Consider the deviation $\tilde{I}_1 = -I_2$. By identical arguments to Case 4 above it must be that $O_{I_1}\left(\tilde{I}_1, I_2, E'\right) \succ O_{I_1}$.[48]

[46] The probability of making the second round is $\frac{1}{2}$ for I_2 both before and after the deviation, so the increase in vote share makes it strictly better off.

[47] As V_{I_2} approaches zero as τ becomes arbitrarily small, I_1 makes the second round for a small enough τ.

[48] Significantly, Cases 4 and 5 covered the possibility that the incumbent parties are able to deter successful entry even if they are located asymmetrically (this possibility will become more significant in the multiple districts section). Implicitly, it was shown that if successful entry can be deterred at asymmetric locations then it can also be deterred if the widest incumbent deviates to symmetry. This follows from Case 4.2.4 that showed if successful entry is possible after this deviation then it must have also been possible before.

2.7.3 Proof of Proposition 2.2: Enter Only if Have a Positive Probability of Victory

The following proof will make use of many of the arguments used in the previous proof (referred to as 'Always Enter'). To avoid repetition such arguments will be referenced here but omitted.

Case 1 $I_1, I_2 \in W'$.

<u>*Case 1.1:*</u> $I_1 = -I_2$.

From 'Always Enter' it must be that $P(W = E) = 0$. This implies $E = \emptyset$ and $P(W = I_1) = P(W = I_2) = \frac{1}{2}$. Consider the deviation $\tilde{I}_1 = I_1 + \alpha$, where $\alpha > 0$. 'Always Enter' Case 1 implies $P(W = E|\tilde{I}_1, I_2, E) = 1 \Longrightarrow P(W = I_1|\tilde{I}_1, I_2, E) = 0$ and therefore $O_{I_1}(\tilde{I}_1, I_2, E') \prec O_{I_1}$. Alternatively, consider the same deviation but where $\alpha < 0$. Lemma 2.1 implies that $P(W = \tilde{I}_1|\tilde{I}_1, I_2, E) = 0$ and again $O_{I_1}(\tilde{I}_1, I_2, E') \prec O_{I_1}$. Thus, $\{I_1, I_2\} = \{y, -y\}$ where $y \in W'$ is a strict Nash equilibrium.

<u>*Case 1.2:*</u> $|I_1| \neq |I_2|$.

By Lemma 2.1, $P(W = I_2) = 0$. If I_2 deviates to $\tilde{I}_2 = -I_1$ then $O_{I_2}(I_1, \tilde{I}_2, E') \succ O_{I_2}$.

Case 2 $I_1 \in W', I_2 \notin W'$.

Identical to Case 1.2 above.

Case 3 $I_1, I_2 \notin \overline{W}$.

In Case 3 of 'Always Enter,' $P(W = E) > 0$ both before and after the deviations considered. Therefore, if the same deviations are considered here it is always the case that $E \neq \emptyset$. As such, the proof here is identical to Case 3 above.

Case 4 $I_1, I_2 \in \overline{W}/W'$.

If in Case 4 from 'Always Enter' both $P(W = E) = P(W = E|E') > 0$ (that is, the entrant has a positive probability of victory both before and after deviations by the incumbents) then the same arguments hold in this case (as the entrant would always enter). This leaves two possible subcases.

Case 4.1: $P(W=E)>0$, and $P(W=E|E')>0$

Therefore, after the deviation studied in 'Always Enter' the entrant can no longer win and stays out (which wasn't allowed in 'Always Enter'). However, the vote share of the incumbents must be as least as great as when the entrant has to enter the market.*[49] As the deviating incumbent's vote share strictly increased in that case, then it must also strictly increase when the entrant chooses to not enter. As the probability of victory was originally zero for all deviators considered in the previous proof, the increased vote share alone implies that the deviator receives a strictly preferred outcome.

Case 4.2: $P(W=E)=0$

Suppose that $|I_1| \leq |I_2|$. If $F(\frac{I_1}{2}) < \frac{1}{3}$ then $E=0 \Longrightarrow P(W=E)=1$. Therefore, $E \neq \emptyset$, which is a contradiction. Thus, only asymmetric incumbent locations when the condition of the equilibrium is satisfied need to be considered (remember symmetric locations when the condition is satisfied may in fact constitute equilibria). Therefore, $|I_1| < |I_2|$ which implies, by Lemma 2.1, that $P(W=I_2)=0$. As the entrant stays out it must be that $E \in (I_1, -I_1) \Longrightarrow V_E < V_{I_1}, V_{I_2}$. Consider then the deviation $\tilde{I}_2 = -I_1$. Obviously it must still be the case that $E' \in (I_1, -I_1) \Longrightarrow V_E < V_{I_1}, V_{I_2}$. As $E' \notin (I_1, -I_1) \Longrightarrow |E'| > |I_1|, \left|\tilde{I}_2\right|$ then $P(W=E|I_1, \tilde{I}_2, E')=0$ and, consequently, $E'=\emptyset$. Thus, it must be that $P(W=I_2|I_1,\tilde{I}_2)=\frac{1}{2}$ and $O_{I_2}(I_1,\tilde{I}_2,E') \succ O_{I_2}$.

Case 5 $I_1 \notin \overline{W}, I_2 \in \overline{W}/W'$.

Proceed as in Case 5 from 'Always Enter', with the same additional remarks as in Case 4 of this proof.

2.7.4 Proofs of the Multiple Districts Results

Proof of Lemma 2.2

[49] And both incumbents are now guaranteed of making the second round. So on this dimension they have to be at least weakly better off.

WOLOG assume that the district in which entry can be deterred has its median at 0 (therefore, $I_1 = -I_2$). Consider now any district i with median at Z_i, where $Z_i \in [-\delta, \delta]$. Let $I_j(i) = I_j + Z_i$ for both $j = 1, 2$. This is possible as $|Z_i| \leq \delta$. As successful entry is deterred in the central district it must also be deterred in district i.

Proof of Lemma 2.3

If the incumbent party platforms are located at I_1 and I_2 then they can choose individual candidate platforms that are symmetric in all districts with medians in the interval $\left[\frac{I_1+I_2}{2} - \delta, \frac{I_1+I_2}{2} + \delta\right]$. As $\left|\overline{Z}\right| = \left|\underline{Z}\right| > \delta$ then it must be that $\left[\underline{Z}, \overline{Z}\right] \nsubseteq \left[\frac{I_1+I_2}{2} - \delta, \frac{I_1+I_2}{2} + \delta\right]$. Thus, there exists at least one interval of district medians such that the incumbent parties can't be located symmetrically. WOLOG, suppose that district j with its median voter at 0 is one such district. I will now establish the conditions such that successful entry can be deterred in this district with asymmetric incumbent locations.

Suppose that in this district $|I_1(j)| < |I_2(j)|$ and consider the outcomes when $E \in I_1^-(j)$ or $E \in I_1^+(j)$.

Case 1: $E(j) \in I_1^+(j)$. This gives the following vote shares: $V_E(j) \to F(\frac{I_1(j)+I_2(j)}{2}) - F(I_1(j))$, $V_{I_1}(j) \to F(I_1(j))$, and $V_{I_2}(j) \to 1 - F(\frac{I_1(j)+I_2(j)}{2})$. As $|E(j)| < |I_1(j)|, |I_2(j)|$, it must be that $V_{I_1}(j) > V_E(j)$ and $V_{I_2}(j) > V_E(j)$. Substituting the above expressions for vote shares and solving gives the following relationships,

$$2F(I_1(j)) \geq F\left(\frac{I_1(j) + I_2(j)}{2}\right) \tag{A}$$

$$1 + F(I_1(j)) \geq 2F\left(\frac{I_1(j) + I_2(j)}{2}\right) \tag{B}$$

Case 2: $E(j) \in I_1^-(j)$. This gives the following vote shares: $V_E(j) \to F(I_1(j))$, $V_{I_1}(j) \to F(\frac{I_1(j)+I_2(j)}{2}) - F(I_1(j))$, and $V_{I_2}(j) = 1 - F(\frac{I_1(j)+I_2(j)}{2})$. As $|I_1(j)| < |E(j)| < |I_2(j)|$ one of the following must be true: (a) $V_{I_1}(j) > V_E(j)$, or (b) $V_E(j) \geq V_{I_1}(j) > V_{I_2}(j)$.

Consider (a) first. This condition implies that $F(\frac{I_1(j)+I_2(j)}{2}) \geq 2F(I_1(j))$. Combining with identity A implies that $2F(I_1(j)) = F(\frac{I_1(j)+I_2(j)}{2})$. Substituting this into identity B produces the additional relationships, $F(I_1(j)) \leq \frac{1}{3}$ and $F(\frac{I_1(j)+I_2(j)}{2}) \leq \frac{2}{3}$. Suppose that $F(I_1(j)) < \frac{1}{3}$. Then it must be that $(\frac{I_1(j)+I_2(j)}{2}) < -I_1(j)$, and it is possible to find a E' such that $(\frac{E'(j)+I_2(j)}{2}) = -I_1$. Then $V_E(j|E') = F(-I_1(j)) - F[-(\frac{I_1(j)+I_2(j)}{2})] = F(\frac{I_1(j)+I_2(j)}{2}) - F(I_1(j))$, and $V_{I_2}(j|E') = 1 - F(-I_1(j)) = F(I_1(j)) = V_E(j|E')$. As $|E'(j)| < |I_1(j)|, |I_2(j)|$ this implies that $P(W = E|E') > 0$ and successful entry is not deterred. So suppose instead that $F(I_1(j)) = \frac{1}{3}$ (and thus $F(\frac{I_1(j)+I_2(j)}{2}) = \frac{2}{3}$) and consider $E' = I_1(j) + \Delta$. Generating the vote shares and substituting the supposition implies that $V_E(j|E') - V_{I_2}(j|E') = 2\int_{\frac{I_1(j)+I_2(j)}{2}}^{\frac{I_1(j)+I_2(j)+\Delta}{2}} f dx - \int_{I_1(j)}^{I_1(j)+\frac{\Delta}{2}} f dx$. By the symmetry of f there exists a Δ small enough such that $V_E(j|E') - V_{I_2}(j|E') > 0$. As $|E'(j)| < |I_1(j)|, |I_2(j)|$ then $P(W = E|E') > 0$ and successful entry is not deterred. Thus, successful entry deterrence consistent with possibility (a) is impossible.

Now consider possibility (b). Firstly, $V_E(j) \geq V_{I_1}(j) \Rightarrow F(I_1(j)) > F(\frac{I_1(j)+I_2(j)}{2}) - F(I_1(j))$. Solving this produces a strict version of condition A. Secondly, $V_{I_1}(j) > V_{I_2}(j) \Rightarrow F(\frac{I_1(j)+I_2(j)}{2}) - F(I_1(j)) \geq 1 - F(\frac{I_1(j)+I_2(j)}{2})$. Solving and combining with condition B implies that $2F(\frac{I_1(j)+I_2(j)}{2}) = 1 + F(I_1(j))$. Applying the strict version of condition A further implies that $F(\frac{I_1(j)+I_2(j)}{2}) > \frac{2}{3}$. This inequality will be used below to establish the lemma. Note that this relationship implies that $F(\frac{0+I_2(j)}{2}) > \frac{2}{3} + \sigma$, for some $\sigma > 0$.

If in district j the incumbents are located asymmetrically such that $P(W = I_2) = 0$ (therefore I_1 is closer to the median) and successful entry is deterred then $I_2(j)$ is aiming to optimize its secondary objectives (i.e., make the second round, maximize vote share). If the entrant enters it maximizes its vote share by choosing $E(j) \in I_1^-(j)$ and $I_2(j)$ is eliminated in the first round. Consider the deviation by $I_2(j)$ to $\tilde{I}_2(j) = I_2(j) - \alpha$, for some $\alpha > 0$. The above analysis implies that the entrant can win the district and that $C^{\varepsilon}_{E(j)}(I_1(j), \tilde{I}_2(j)) = I_1^-$. Further, $I_1(j)$ is eliminated in the first round and thus $I_2(j)$ must survive until the second round. Therefore, as $V_{I_2(j)}(I_1(j), \tilde{I}_2(j)) = 1 - F(\frac{I_1(j)+I_2(j)-\alpha}{2}) > 1 - F(\frac{I_1(j)+I_2(j)}{2}) = V_{I_2(j)}$, candidate $I_2(j)$

is strictly better off. Consequently, in all districts j in which $P(W = I_2) = 0$ it must be that $I_2(j) = I_2 - \delta$ (i.e., this candidate can't move any further to the left).

If in all districts $P(W = I_2) = 0$, and successful entry is deterred, then consider the deviation of the party platform $\tilde{I}_2 = I_2 - \alpha$, for some $\alpha > 0$, and the candidate platforms, for all j, $I_2(j) = \tilde{I}_2 - \delta$. By the above analysis this implies that party I_2 is strictly better off in every district and thus $O_{I_2}(I_1, \tilde{I}_2) \succ O_{I_2}$. Therefore, there must exist an interval of districts $(\overline{s}, \overline{s} + \omega)$ such that $\forall Z_l \in [\overline{s}, \overline{s} + \omega]$, $P(W = I_2(l)) > 0$ (and at least one of these must require $I_2(k) = I_2 + \delta$ otherwise the above deviation of the party platform would still be applicable).

Assume, without loss of generality, that $\overline{s} = 0$ (and the interval is $(0, \omega)$). Further, consider the case when successful entry is deterred in the district with its median voter at 0 (if instead entry is deterred in favor of I_1 in the district with its median at ω then analogous analysis holds with greater with even greater vote shares for the entrant). The analysis above showed that $F(\frac{0+I_2(j)}{2}) > \frac{2}{3} + \sigma$, for some $\sigma > 0$. As $P(W = I_2) > 0$ in all districts k such that $Z_k \in (0, \omega)$, Lemma 2.1 implies that $|I_2(k) - Z_k| \leq |I_1(k) - Z_k|$. Combining these relationships, and assuming $E(k) = 0$, implies that $V_{E(k)} = F(\frac{0+I_2(k)}{2}) - F(\frac{I_1(k)+0}{2}) > \frac{1}{3}$ for small enough ω. As $|E(k)| < |I_1(k)|, |I_2(k)|$ it must be that $P(W = E) = 1$ in these districts. This contradicts the requirement that there exist some districts in which $P(W = I_2) > 0$ and, therefore, proves that an equilibrium that deters successful entry in all districts cannot exist when $|\overline{Z}| = |\underline{Z}| > \delta$.

Proof of Proposition 2.3

Firstly, as the incumbent location pair $\{I_1^*, -I_1^*\}$ deters entry in the single district game, it must be the case that all location pairs $\{I_1, -I_1\}$ such that $I_1 \in [I_1^*, 0]$ also deter successful entry in the single district game. This follows as successful entry on the flank is still not possible (as $|E| > |I_1|, |I_2|$) and for any $E \in (I_1, -I_1)$ it must be that V_E is strictly lower and V_{I_1} and V_{I_2} strictly higher than for the same E given the pair $\{I_1^*, -I_1^*\}$. Therefore, as successful entry is deterred for the pair $\{I_1^*, -I_1^*\}$, it must also be deterred for the pair $\{I_1, -I_1\}$ if $I_1 \in [I_1^*, 0]$.

For district i, $I_2(i) - Z_i = \tilde{I}_2 + \epsilon \leq I_2^*$. As $|I_1(i) - Z_i| = |I_2(i) - Z_i| \leq I_1^*$ it follows that successful entry is deterred in district i and every district is in equilibrium. As there exists no profitable deviations in any district (even without the restriction on candidate movement), there exists no profitable deviations for the party. Therefore these strategies constitute an equilibrium.

For the second half of the proposition consider the following set of strategies: $\{I_1, I_2\} = \{\tilde{I}_1, \tilde{I}_2 + \eta\}$, and $I_1(i) = \tilde{I}_1 + Z_i, I_2(i) = \tilde{I}_2 + Z_i - \eta$. If $\eta < \delta - |\underline{Z}|$ then these candidate positions satisfy Assumption 2.7 . As above, there are no profitable deviations in each district. Therefore, there are no profitable party deviations and this is an equilibrium.

Part II

Vote Timing and Information Aggregation

Introduction

In contrast to the candidate centered approach of Part I, Part II concentrates on the incentives and strategic decisions of the voters themselves. Unfortunately, the environments in which real elections are held, and vote decisions made, are typically far from ideal. Voters are often required to make their decisions with only limited information about the preferences and characteristics of both the competing candidates and their fellow voters. As such, it is not clear that the voting mechanism will lead to acceptable outcomes. The literature on voting theory, to which the following two chapters belong, is an attempt to examine this very possibility.

The standard way of modelling the voting problem is to assume that the candidates differ on both quality and policy dimensions. Voters are assumed to be aware of the candidates' policy platforms, but only imperfectly informed about the relative quality of the candidates. The problem then for the voters is that they are not sure which candidate, if elected, would provide them with the maximum benefit. The question of how voters respond to this situation, and the outcomes that are produced from this behavior, has been the focus of the voting literature, and is the focus of Part II of this thesis.

Until relatively recently this voting decision was not thought to require much analysis. Each voter, quite simply, was expected to cast a ballot in accordance with the private information she possessed (after allowing for her personal preferences). However, in a series of important contributions, Lohmann (1994), Austen-Smith and Banks (1996), and Feddersen and Pesendorfer (1997) showed that such behavior was not always rational, and, in fact, that rational behavior required much greater strategic analysis than previously thought. Basically, these papers showed that the voting situation described above constituted a game, and therefore the voters should be allowing for, and reacting to, the behavior of their fellow voters.

The key insight of this research is that voters should condition their actions on being pivotal as this is the only situation in which their action will affect their utility. Perhaps surprisingly, it was then observed that such conditioning does not always lead to voters casting ballots in accordance with their private information. Consequently, it may be that not all relevant and valuable information is being used in deciding election outcomes.

Given this seminal observation, research has focused on the question of how effective voting mechanisms are in aggregating dispersed private information in the making of group decisions. The benchmark point of comparison is typically taken to be the outcomes that would result if all information was public knowledge. In common interest environments this equates to the set of Pareto optimal outcomes. Previous work, most notably by Feddersen and Pesendorfer (1997), has suggested that in simultaneous elections the plurality rule mechanism does, despite the conflicting objectives and dispersed information, achieve optimal outcomes in large electorates.

In Chapters 3 and 4 I investigate the effect of the timing of votes on this capability. For clarity I study a common interest environment and attempt to explain how the timing of votes can affect the behavior of voters, the outcomes that result, and the quality of information aggregation within the plurality rule voting mechanism. In Chapter 3 I firstly study a model of sequential voting and explain when and why the commonly observed phenomena of bandwagons and momentum arise. I show that only if voters have a desire to vote for the winning candidate, in addition to their desire to select the better quality candidate, is momentum observed and bandwagons begin. In Chapter 4 I go on to compare these results with analogous results for when voting is simultaneous and characterize when each process is superior in aggregating information and producing efficient outcomes. I find that in lopsided races, when a strong front runner exists, sequential voting is preferred, but in tight races, when such a front runner doesn't exist, simultaneous voting is better. Strangely, the superior performance of sequential voting in lopsided races is precisely because bandwagons occur.

For these two chapters I am able to achieve results only at the expense of assuming there is a countably infinite number of voters. As infinite voting populations are rarely observed, I also examine in Chapter 4 a model of simultaneous voting when voters have a desire to vote for the winner and when the voting population is finite.

Reassuringly, I find that the results for large populations are well approximated by the infinite population limit results.

Chapter 3 Bandwagons and Momentum in Sequential Voting

Abstract

In this chapter I show that an equilibrium exists to the sequential voting game in which a bandwagon begins with probability one. These bandwagons are driven by a combination of beliefs and the desire of voters to vote for the winning candidate. Significantly, in this equilibrium the pivot probability for each voter is nonzero, even in an infinite population. Consequently the bandwagons do not always start after one (or at most two) favorable decisions (as do economic cascades) and varying levels of informative voting are observed, consistent with observations from sequential voting in U.S. presidential primaries. Further, voters are exposed to counterintuitive incentives, referred to as "buyers' remorse," that have been attributed to real primary voters.

From the play of this equilibrium an explanation of momentum arises that is consistent with empirical regularities. This interpretation provides a formal distinction between the often ambiguous concepts of momentum and bandwagons, and permits a separation of their effects on the sequential voting mechanism. Finally, I observe that the limiting cases of the generalized bandwagon voting strategy (BWV) employed are informative and uninformative voting. This not only uncovers a natural relationship between these three (previously thought to be distinct) voting strategies, but also provides a connection between the positive results presented here and previous negative findings on bandwagons in sequential voting.

3.1 Introduction

The emphasis placed on "momentum" and "bandwagons" during the American presidential primary process might well overwhelm even the most casual observer. Both the media and the voting population pay close attention to the current status of "momentum" and the possible onset of a "bandwagon." Even the candidates themselves believe in the power of these phenomena.[1] In response, an extensive academic literature has arisen that has not only confirmed the existence of these phenomena, but in fact concluded that they can alter the outcome of an election.[2] Remarkably, despite this emphasis and apparent importance, there does not exist a rigorous theoretical explanation for these phenomena.

In fact, there does not even exist standardized definitions of the dynamic forces these terms are intended to describe. There are common underlying themes, however, and these are captured by the following informal descriptions: a "bandwagon" is said to have begun if a point is reached such that (nearly) all subsequent voters vote for the leading candidate, and a candidate is said to have "momentum" when his chances of victory are improving.

Though the presidential primaries are the most visible (and studied) example of sequential voting, the questions raised here are applicable to any decision situation in which a voting process is used. It is possible for a "bandwagon" to start for an investment option when votes are cast around a boardroom table just as one often starts for fortunate presidential candidates. Consequently, the dynamic phenomena of "bandwagons" and "momentum" impact all aspects of society where voting processes are employed. These settings are so diverse as to include shareholders voting at annual

[1] An explicit example of this is found in an interview with George Bush after his success in the 1980 Iowa caucuses (Greenfield (1982, p.39-40)),

> "What we'll have, you see, is momentum. We will have forward 'Big Mo' on our side, as they say in athletics."
>
> "'Big Mo'?," Schieffer asked.
>
> "Yeah," Bush replied, "'Mo,' momentum."

In addition to this, there are many other, less explicit, examples of candidate manipulation and strategy aimed at taking advantage of these phenomena (see Bartels (1988)).

[2] Jimmy Carter in 1976 is the most commonly cited example (see Bartels (1988)).

meetings, the formulation of supreme court rulings and even (in principle) the setting of monetary policy by central banks.

Given the significance of the decisions that are made, or could be made, with sequential voting it would seem an important objective to understand why momentum is observed, and when and why bandwagons begin. The answers to these questions can explain how momentum and bandwagons affect the selection of candidates via sequential choice, and whether the winning candidates are the "best" choices available to the electorate. In this chapter I attempt to answer these very questions. I present a model of incomplete and asymmetric information in sequential voting in which both momentum and bandwagons are observed. Further, I characterize the conditions that lead to the start of a bandwagon and quantify the cost of this phenomenon to society in terms of the quality of candidate that is selected. In Chapter 4, I continue this work and compare the information aggregation properties of the sequential voting mechanism unearthed here with those of the simultaneous voting mechanism.

Previously it had been thought that the models of informational cascades from the economics literature could be applied directly to models of sequential voting.[3] In an informational cascade, agents ignore their private information and instead mimic the behavior of previous agents. However, the models of cascades in economics are of individual action, not collective choice. Put simply, the economic agent consumes her choice, regardless of the choices of others, whereas the voter consumes the group selection independent of her own choice. Consequently, voters must concern themselves not only with the choices made before them, but also with how their choice will influence the choices of voters following them.

Fey (1998) and Wit (1997) show that these differences are more than trivial. They show that while an economic type of informational cascade can be supported as a perfect Bayesian equilibrium in a sequential voting model, this equilibrium is only supportable by severely restricting the inferences later voters can make about the actions of earlier voters. Further, once a "sensible" (Fey (1998)) restriction on beliefs (possible inferences) is imposed, an informational cascade cannot be supported as an

[3]Bikhchandani et al. (1992) cite voting as an application for their results.

equilibrium. Instead, Fey and Wit find that "informative voting" (meaning all voters vote according to their private information) is a perfect Bayesian equilibrium to the voting model, and that this is true without any restriction on inferences.

These negative results do not correspond with the empirical observation of momentum and bandwagons in sequential elections. They do, however, show that voting bandwagons must be caused by forces other than those causing economic cascades. In this chapter I derive positive results supporting bandwagons and momentum. The bandwagons I find, unlike economic cascades, are not driven solely by individuals' beliefs but also by the desire of voters to vote for the winning candidate.

This different cause results in bandwagons that are much more general than are economic cascades. Voters will "jump on the bandwagon" when the support for the leading candidate passes through a certain threshold. This threshold is flexible: it can be so low that a bandwagon starts immediately, or so high that no matter how much support a particular candidate receives he will never benefit from a bandwagon. In contrast, economic cascades do not exhibit this flexibility, and instead have a fixed threshold. The flexibility in these voting bandwagons is consistent with the different types of bandwagons observed in sequential elections.[4]

The bandwagon threshold is determined by the degree of the voters' desire to be on the winning team. For any particular intensity of this desire it is possible that several different threshold levels are supportable as equilibria. As this willingness approaches zero the minimum number of votes required before the bandwagon begins approaches infinity. However, the key finding is that as long as the desire to vote for the winner is nonzero, a bandwagon will ultimately occur.

In the limiting case, when voters gain no additional utility from voting for the winner, the threshold becomes so high that it can never be reached, implying that all voters vote "informatively" and a bandwagon never starts! This is the case considered by Fey (1998) and Wit (1997), thus their negative results can be seen as the special limiting case of the positive results presented here. At the other extreme, if the utility gain from voting for the winner is sufficiently high then in equilibrium all voters will

[4]Bartels (1985).

disregard their private information and "bandwagon" immediately. This behavior is sometimes called "uninformative voting." These limit results uncover a natural relationship between "informative," "uninformative," and "bandwagon" voting in which these three, previously thought to be independent, voting strategies can be seen as special cases of the one general voting strategy.[5]

From the features of the bandwagon equilibrium just described, an explanation of momentum arises quite naturally. This interpretation of momentum displays characteristics that are consistent with the most commonly cited empirical regularities: that primary election wins tend to lead to greater voter support and improved performance in subsequent primaries and, most critically, that momentum can be reversed. I will show that these characteristics arise as mechanical by-products of the model of incomplete information and Bayesian updating employed here. This interpretation provides a formal separation and distinction between the often ambiguous concepts of momentum and bandwagons in sequential elections.

To support these results, one can cite many possible reasons why a voter may gain additional utility from voting for the winning candidate. These range from the pure psychological benefits of being on the winning team or conforming with the majority, to the benefits of rational strategic considerations. The behavior of "yes men," (people who always agree with another's suggestion) is consistent with such a desire. A more overt example is the people of New Hampshire, who enjoy voting for the ultimate winner of the primary contests to such an extent that they boast about their ability to regularly do so, such as via a sign in Manchester airport that reads, "Always First, Always Right."[6] Confirming this intuition, Niemi and Bartels (1984) and Bartels (1988) provide rigorous experimental and empirical evidence, respectively, consistent with such a desire. They show that a voter's likelihood of voting for a candidate increases with her belief that this candidate will win, and that this impact is over and above the effect the belief has on her underlying preferences.

This extension formalizes and makes explicit the notion that voters derive an in-

[5]As defined by Fey (1998).

[6]As described in the Los Angeles Times (Fiore (1/26/00, p. A1)).

trinsic benefit from casting a ballot. Such a notion is absent from traditional rational choice models, causing them to be plagued by the "paradox of turnout." This conundrum is that if voting is costly (in time and effort etc.), and voters have the option to abstain, then in any rational choice equilibrium turnout is far lower than is empirically observed. This paradox remains an open question, but is typically ignored under the (usually unstated) belief that at least some voters derive an intangible benefit from voting. In this chapter I convert this implicit belief into an explicit assumption and show that if the benefit of voting depends on expectations as to which candidate will win the election, as is empirically documented, then bandwagons and momentum will be observed in sequential elections.

The primary contribution of this chapter is to provide a possible explanation to an open question: why are bandwagons and momentum observed in sequential voting? Whilst this explanation, that voters have a desire to vote for the winner, is seemingly intuitive, the voting behavior and equilibrium that arise from this incentive are in fact rather surprising when we consider the problem more deeply. The standard intuition from voting games is that the ability of any single voter to make a difference, his "pivot probability," approaches zero as the population becomes large. Consequently, the marginal utility for a voter of using private information to advantageously affect the outcome disappears in the limit. With the infinite population assumed here it may therefore be expected that the addition of a desire to vote for the winner would swamp the informational incentive and result not only in bandwagons, but bandwagons that start immediately. Obviously, such bandwagons would not resemble the flexible bandwagons described above that are found in real sequential elections. The flexible bandwagons found here may also be surprising from the opposite perspective. Alternative intuition may be that the inclusion of a desire to vote for the winner would not make any difference at all to the models of Fey (1998) and Wit (1997) given the game remains one of common interest (as ideally all voters would still like to make the same correct choice).

Interestingly, I find neither of these extremes to be the case and instead find that pivot probabilities do not go to zero for any voter. Consequently, my model predicts a combination of informative and uninformative voting that is consistent with the empirical phenomena of bandwagons and momentum.

In addition to this, I uncover unexpected incentives facing voters both before and after a bandwagon begins. I find that, once a candidate has achieved more than a two vote lead, all voters wish to vote against this leader in order to increase the probability of selecting the best candidate, even those voters whose private information indicates that he is the best candidate. Surprisingly, it is only the desire to vote for the winner that causes these voters to support the leader. These incentives are surprising as informational incentives alone typically lead the voter to cast his ballot informatively when other voters are behaving symmetrically, as was found to be the case in the similar models of Fey (1998) and Wit (1997). Significantly, these strange incentives have actually been observed in presidential primaries and are referred to as "buyers' remorse."[7] Similar to the incentives faced by voters in my model, "buyers' remorse" describes the hesitation of voters to support the leading candidate for fear that it would end the race prematurely.

These incentives also show that the existence of bandwagons with flexible thresholds is more than an issue of beliefs, as was the case with the limited cascades of Fey (1998) and Wit (1997). In fact, regardless of beliefs, voters must gain additional utility from voting for the winner to make following the bandwagon their optimal action, even when the equilibrium requires them to vote "informatively" before the bandwagon has begun.

Taken together, these counterintuitive incentives and characteristics add up to the intuitive solution that a desire to vote for the winner leads to the onset of bandwagons and momentum. Additionally, these findings indicate the hurdles other potential explanations of bandwagons would need to overcome and the characteristics they must exhibit.

The remainder of the chapter is organized as follows. The following two sections present the model and the results. In section 4, I motivate and justify the assumption that voters derive additional utility from voting for the winner, and discuss the nature

[7] Los Angeles Times (Brownstein (1/26/00, p. A8)).

of the bandwagons and momentum found in the model. The final section concludes and suggests several directions for future research.

3.2 The Model

The model is one of sequential voting with incomplete information. There are a (countably) infinite number of voters who cast ballots sequentially, in a fixed order.[8] The voters are labeled 1,2,3,...,i,... . There are two candidates, A and B. The candidate who wins is determined according to majority rule. The majority rule for an infinite population is defined below. This definition simply claims that if one candidate forges a vote share lead that he never relinquishes, even in the limit, then he is declared the winner. If neither candidate can achieve such a lead then the election is declared a tie and the winner is decided by the toss of a fair coin. When the voting population is finite this definition collapses to the standard definition of majority rule. Let v_i denote the vote cast by voter i, and W the winner of the election. The vote share of candidate A after n votes have been cast is $S_n(A) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{v_i=A\}}$.

Definition 3.1 *The majority rule winner is,*

$$\begin{aligned} W &= A \text{ if } \liminf S_n(A) > \frac{1}{2}, \\ W &= B \text{ if } \limsup S_n(A) < \frac{1}{2}, \text{ or,} \\ P(W=A) &= P(W=B) = \frac{1}{2} \text{ otherwise.}^9 \end{aligned}$$

[8]The assumption of an infinite number of voters is, of course, an approximation to reality. It is in the same spirit as the assumption of a continuum of voters in simultaneous voting games and aids considerably in the tractability of the problem. In an economic context this assumption was made by Lee (1993) in an analysis of the convergence characteristics of cascades. One additional problem that arises due to the sequential voting context is with respect to the duration of the election. If there is equal temporal spacing between voters then either the election will go on forever or there is a nonmeasurable amount of time between each voter. However, a simple solution is provided to this dilemma by a recourse to the basic properties of a convergent geometric progression. If instead the time between voters is decreasing by a factor less than one then the election is finished in finite time and there is still a measurable amount of time between any pair of voters.

[9]As $S_n(A)$ is a bounded sequence the lim inf and the lim sup must exist. It is easy to see that the winner under this rule is invariant to the reordering of any finite set of voters. However, given

There are two possible states of the world, also labeled A and B. Each voter has identical preferences dependent upon whether the correct candidate is chosen (A in state A, and B in state B) and whether they vote for the winning candidate (reward of $k \geq 0$).[10,11] If U_i denotes voter utility, then these preferences for an arbitrary voter i can be represented (loosely) by,

$$U_i = 1_{\{(W=A|A),(W=B|B)\}} + k.1_{\{v_i=W\}}$$

This utility function is more general than that typically used in models of voting (e.g., Austen-Smith and Banks (1996), Fey (1998)). The extension is the inclusion of the possibility that voters receive additional utility if they vote for the winning candidate. Previous models considered only the limit case of $k = 0$. Note that despite the difference between this utility specification and previous models, the common values component implies that if information was complete all voters would still ideally wish to make the same decision.

the nature of countably infinite sequences this is not necessarily true for reorderings of infinite sets of voters. Though, given the different sequences resulting from such a rearrangement, it would not seem that invariance to such rearrangements is in fact a desirable property of an electoral rule.

There are many alternative ways that majority rule can be defined for a countably infinite voting population, each with its own shortcomings. This definition is used for several reasons: it is equivalent to majority rule when the population is finite, it captures the idea that the ability of a single voter to affect the outcome disappears in the limit, and it is unaffected by simple changes in the order of vote counting that afflict other possible definitions. For example, an obvious alternative is that a candidate is declared the winner if it forges a vote lead (vote count and not vote share) that never disappears. Consider the outcome then when the first person votes A and after that all men vote A and women vote B. If counting alternates between the sexes then the outcome changes depending on whether we start with a man or a woman (AABABAB⇒A wins, versus ABABABA⇒ tie), an obviously undesirable feature. The definition employed here is impervious to this type of rearrangement. It should be noted though that regardless of how the technicalities in close elections are dealt with, the results presented here are unaffected as in all of the equilibria the vote share of the candidate deemed victorious approaches one, and thus should be selected by any reasonable counting rule.

[10] As mentioned in the introduction of this chapter, and discussed in Section 3.4.1, there exists ample motivations for why a voter would receive additional utility by voting for the candidate most likely to win the nomination, as well as extensive empirical and experimental evidence that voters do indeed vote with these incentives in mind.

[11] It should be noted that the results presented here are not dependent upon all voters having identical k. The results hold as stated if instead it is only required that all k_i are within some neighborhood of a common k. The robustness of the model to greater heterogeneity will be explored on page 129.

Each voter receives an independent, private signal about the true state of the world. This signal is either α or β and is sent accurately with some probability $p > \frac{1}{2}$. That is, $P(\alpha|A) = P(\beta|B) = p$. The state of the world can be interpreted as specifying which candidate is unambiguously "better" in the minds of the voters. The voters share a common prior, π, that the true state is A. I assume priors are "permissible," which is given by the following definition.[12]

Definition 3.2 *Priors are "permissible" if there exists an integer m such that the following expression holds,*

$$\pi = \frac{p^m}{p^m + (1-p)^m} \tag{3.1}$$

This is a restriction on permitted priors that is not usually made. It is made here in order to permit tractability. The proof of the main result employs this restriction but at no point does it appear critical. It will be seen later that the bandwagon equilibria are robust to at least small perturbations of these priors. Priors will be said to be "neutral" when $m = 0$ ($\pi = \frac{1}{2}$).

As the voters cast their ballots in sequence there develops a publicly observed history of votes. For any voter i the relevant history, denoted $h_i = (v_1, v_2, ..., v_{i-1})$, is the vector of votes cast when it is her turn to decide. For the first voter this history is empty.[13] Denote by s_i the private signal of voter i. A *strategy* for voter i can then be defined as a map $\sigma_i : (s_i, h_i) \to [0, 1]$ describing the probability i votes for candidate A as a function of her information: the expressions $\sigma_i\,(A|s_i, h_i)$ and $\sigma_i\,(B|s_i, h_i)$ denote the probabilities of voting for candidates A and B, respectively, given this information. A *voting profile* is then denoted by $\sigma = (\sigma_1, \sigma_2, ..., \sigma_i, ...)$. Let σ_{-i} represent the voting profile for all voters other than i.

After any history, h_l, define voter j's conditional probability that voter k's signal is s_k to be $\mu_j(s_k|s_j, h_l)$. Denote the belief profile by μ. Bayes' Rule is used to

[12]It may seem odd that the prior beliefs permitted depend on the precision of the voters' private signals. A possible justification for this is that before any of the private signals are revealed there are an arbitrary number of signals that are publicly revealed, and that m more of these favored candidate A rather than candidate B.

[13]If voting is simultaneous, as will be considered in Chapter 4, then this history is empty for all voters.

update beliefs after all v_i when $\sigma_i(v_i|s_i, h_i) > 0$ for some s_i. Expected utility for voter i, u_i, can now be defined formally as a map $u_i : (\sigma, s_i, h_i) \to \Re$. A Perfect Bayesian Equilibrium (PBE) is a voting profile, σ, and a belief profile, μ, such that $u_i(\sigma|s_i, h_i) \geq u_i(\sigma_i', \sigma_{-i}|s_i, h_i)\ \forall i, \sigma_i', s_i, h_i$.[14]

Given that a voter's strategy may depend on her private signal, her vote choice may contain some of this private information. The other voters will use this information to update their beliefs as to which is the best candidate. For the history h_i denote the belief of voter i that A is the best candidate by $\pi(h_i|\pi, \sigma, \mu)$. In making a decision a voter combines this belief with her private signal and forms the updated belief, $\varphi(h_i, s_i|\pi, \sigma, \mu)$. For ease of exposition the triple, (π, σ, μ), will be omitted from the notation for the remainder of the chapter.

Voting Strategies

Both the beliefs held and the private signal observed by a voter affect the expected utility from each decision choice. One possibility is that the signal will dominate these calculations and consequently the voter will always vote in agreement with it. Alternatively, the beliefs may dominate if they reach a certain level, or a combination of the two factors may influence the vote cast. The following general Cut Point Voting strategy incorporates all of these possibilities.[15]

Definition 3.3 *Cut Point Voting (CPV) is a strategy, σ_i, such that*

$$v_i = \left\{ \begin{array}{c} B \text{ if } \pi(h_i) < C_B \\ A \text{ if } \pi(h_i) > C_A \\ s_i \text{ otherwise} \end{array} \right\}$$

Where $C_A, C_B \in [0,1]$ and $C_A \geq C_B$.

Denote by CPV(C_B,C_A) the strategy when the thresholds are C_B and C_A. As the parameters C_A and C_B vary the behavior of a voter using a CPV strategy also

[14] If voting is simultaneous then this definition is equivalent to a Nash equilibrium.

[15] Though $s_i \notin \{A, B\}$, I abuse notation and use $v_i = s_i$ to represent the action of voting informatively (i.e., voting in accordance with the private signal).

varies. I will be particularly interested in certain subsets of this generalized strategy. If either C_A,$C_B \notin \{0, 1\}$ then each voter "votes her signal" until beliefs pass a given threshold in favor of either candidate, at which point the subsequent voters "jump on the bandwagon" and vote for the leading candidate regardless of their private signal. I will call this set of strategies Bandwagon Voting and denote it by BWV(C_B,C_A).

Given this strategy, a bandwagon is the vote sequence after a threshold is passed. Thus a bandwagon is said to have commenced when all voters begin voting for the leading candidate regardless of their private signals, and a victory for this candidate becomes inevitable.[16] It is easy to show that if both C_A,$C_B \in (0, 1)$ then a bandwagon will start with probability one.[17]

This definition of bandwagon voting differs from the definition of cascades used both in the economics literature (e.g., Bikhchandani, et al. (1992)) and voting models (Fey (1998), Wit (1997)). Cascades restrict attention to the case where $C_A = p$ and $C_B = 1 - p$. Consequently, depending on the exact value of π, once either a one vote or, at most, two vote lead is established by a candidate the contest is effectively over as all subsequent voters cascade onto this candidate.

The bandwagon behavior found in my model is of a far richer, and more realistic, variety. Depending on priors and the value of the parameters, k and p, in equilibrium the bandwagon may require a vote lead of any amount before a candidate benefits from its effects. Though previous attempts at theoretical descriptions of this phenomenon have not allowed for this generality, it is precisely this type of bandwagon threshold that meshes well with empirical observation. This is made clear by Bartels (1985),

> "Rather than doing better and better (or worse and worse) in an unbroken cycle, candidates may reach plateaus of support determined in part

[16]This is consistent with the description by Berelson, Lazarsfeld, and McPhee (1954, p.289), "*... a "bandwagon effect"; people may vote for the man whom they expect to be the winner.*" Though in the literature there are many, often contradictory, descriptions of bandwagons and the situations where they may occur, the notion described here seems to capture the common theme that voters will support a candidate simply because he is expected to win.

[17]This is proven by showing that the probability a bandwagon hasn't started is bounded above, and that this bound approaches zero as the number of voters who have cast their ballots approaches infinity. Therefore, with probability one a bandwagon must eventually start in finite time. This is shown formally in the appendix.

by their political skills and circumstances." (p.814)

The generalized CPV definition also contains as special limit cases two other voting strategies that were previously considered distinct from bandwagon voting and defined separately. The first is "informative voting," which is when a voter votes her signal regardless of other information. This corresponds to CPV(0,1).[18] Likewise, "uninformative voting," when a voter selects a candidate independent of her private signal, corresponds to CPV(0,0) when picking candidate A, and CPV(1,1) when picking candidate B.[19] Therefore, instead of considering these different voting strategies as competing theories of sequential voting and trying to decide between them, they should be thought of as special cases of the one general theory.

In this light many apparently different nomination battles can be seen as consistent with the one CPV strategy, and instead are simply examples of different parameter values. Specifically, "informative voting" is the limit case of BWV in which the bandwagon threshold is never reached.

I will also use these same terms to describe certain actions. I will say that, regardless of the strategy being employed, a voter voted "informatively" if her vote choice conveys her private information to the other voters. Likewise, she will be said to have voted "uninformatively" if her vote choice reveals no information. For example, all voters using a BWV strategy are required to vote "informatively" until a threshold is passed, after which they vote "uninformatively." Further, if a BWV strategy is being played by all voters in equilibrium then I will refer to this as a BWV equilibrium. Analogous terminology will be used for "informative voting," "uninformative voting," and CPV equilibria.

The definition of a PBE stated earlier places no restriction on the beliefs that can be formed after a zero probability event is observed. The freedom of this variable in games of asymmetric information often leads to the problem of multiple equilibria.

[18] As it is assumed that m is finite it must be that $\pi \notin \{0,1\}$.

[19] If all voters are using a BWV strategy and $\pi \geq C_A$, then as long as there are no deviations all voters will vote uninformatively for candidate A. Observationally this path would be equivalent to all voters using an "uninformative voting" strategy. However, these two strategies are not equivalent for $C_A > 0$ as for BWV there exists a possible sequence of deviations (and beliefs) following which a voter using the BWV strategy would have to vote for B.

It is argued by Fey (1998) and Wit (1997) that in this voting context only a certain refinement of these beliefs should be considered, what Fey calls a "sensible" refinement.[20] Their argument is that if a voter who is expected to vote uninformatively for A actually votes for B then subsequent voters should believe with probability one that this voter observed a signal of β. That is, they should not expect a voter to deviate from a bandwagon if she observed a signal supporting the leading candidate. As Fey observed, this reasoning is similar to that of the Intuitive Criterion (Cho and Kreps (1987)). In the language of game theory, any deviation is "maximally informative." This reasoning is even more appropriate with the utility function employed here. I will refer to beliefs that satisfy this refinement as "sensible." Formally this restriction is the following.

Definition 3.4 *Consider any pair of voters i and j where $j < i$, and v_j, where $\sigma_j(v_j|s_j, h_j) = 0$ for all $s_j \in \{\alpha, \beta\}$. Beliefs are considered to be "sensible" if,*

$$\mu_i(s_j = v_j|s_i, h_i) = 1$$

The arguments for this refinement are persuasive and it shall be employed throughout the chapter.[21] Once "sensible" beliefs are imposed the difference between the model presented here and previous work becomes clear. In applying economic cascades to the voting model, Fey (1998) and Wit (1997) find that a cascade can be supported as an equilibrium, but not once beliefs are required to be "sensible."[22] Instead they find that "informative voting" is supportable as an equilibrium for any

[20]This refinement does not impact the analysis when voting is simultaneous.

[21]For the restricted cascades considered by Fey, equilibria without "sensible" beliefs can be eliminated by a weak version of the Intuitive Criterion (make inequality signs weak). However, for the general bandwagons considered here such a refinement is not possible, and instead arguments more in the spirit of the Divinity Criteria are required (Banks and Sobel (1987)). Though, significantly, for $k > 0$ this refinement of beliefs is not necessary for any of the results. Regardless of beliefs a bandwagon equilibrium exists, voter behavior on the equilibrium path is unchanged, and "informative voting" is not an equilibrium.

[22]Essentially the beliefs which support cascade behavior as an equilibrium require that any deviation from the cascade be ignored by later voters.

specification of beliefs.[23]

These negative findings, which at first glance appear fatal to the prospect of bandwagons, disguise the subtlety of each voter's decision process. Significantly, "informative voting" is not only the limiting case of CPV, but it is the only case for which a bandwagon doesn't start. Therefore, in the "informative voting" equilibrium found by Fey and Wit for the case of $k = 0$ voters would be willing to "jump on the bandwagon," but beliefs never quite reach the threshold for this to happen.

This interpretation implies that bandwagon behavior is not as unattractive to voters as previously thought. In fact, later voters approach indifference between voting informatively and bandwagoning, though always marginally preferring the former. In this chapter I show that if voters receive additional utility by voting for the winning candidate then this ordering is reversed and voters strictly prefer to bandwagon than vote informatively once either candidate achieves a large enough vote lead. This is true for even an arbitrarily small k. Therefore, in equilibrium the CPV threshold is lowered and a bandwagon will start with probability one. In fact, as k approaches zero the minimum number of votes required before a bandwagon starts approaches infinity, which is when a bandwagon never starts. Thus, the findings of Fey and Wit can be seen as a limiting case, and the only case in which a bandwagon will never start, for when voters derive utility from voting for the winner.

If all voters are using the BWV strategy and beliefs are "sensible" then behavior can be described as follows: all voters vote their private signals until beliefs in favor of one of the candidates passes through a threshold, at which point all subsequent voters vote for this candidate regardless of their private signals. For beliefs to reach the threshold a candidate requires a certain "vote lead," which is the number of votes for one candidate in excess of the number of votes for the other candidate. By the properties of Bayesian updating with equally precise signals the "vote lead" is, up until the bandwagon starts, a sufficient statistic for the information contained in the vote history. Thus, in order to adhere to the BWV strategy each voter need only

[23]As all histories are reached with positive probability if voting is informative, all beliefs are determined by Bayes' rule.

know their private signal and the vote lead.

However, when beliefs are "sensible," a difficulty arises in that after the bandwagon starts some votes contain no information whereas others fully reveal the voter's private signal (when a voter deviates from the bandwagon). In order to keep track of the information contained in votes both before and after the bandwagon has started, I will define the "informative vote lead." This measure is similar to the "vote lead," but instead ignores votes that reveal no private information. This measure, denoted by n_{i+1} when i votes have been cast, is positive when A has received more "informative" votes than candidate B, and negative otherwise. Assuming that all voters are using the BWV strategy, n_i can be defined as follows.

Definition 3.5 *The "informative vote lead" faced by voter i is,*

$$n_i = \sum_{j:\pi(h_j)\neq\pi(h_{j+1})} \left[1_{\{v_j=A\}} - 1_{\{v_j=B\}}\right] \tag{3.2}$$

With an infinite number of ballots to follow each voter, voters k and l face the same decision problem if $n_k = n_l$. Thus, to simplify notation I will omit the subscript on n. As n is a sufficient statistic for the history of votes it can be used to simplify the expressions representing the beliefs of voters. Note that these expressions hold even if a voter should deviate after the onset of a bandwagon.

$$\begin{aligned}
\pi(n) &= \frac{p^{m+n}}{p^{m+n} + (1-p)^{m+n}}, \\
\varphi(n,\alpha) &= \frac{p^{m+n+1}}{p^{m+n+1} + (1-p)^{m+n+1}}, \text{ and,} \\
\varphi(n,\beta) &= \frac{p^{m+n-1}}{p^{m+n-1} + (1-p)^{m+n-1}}
\end{aligned}$$

Obviously, $\varphi(n-1,\alpha) = \varphi(n+1,\beta)$. To simplify the notation further these arguments shall be combined in the expression of φ such that $\varphi(n+1) = \varphi(n,\alpha)$ and $\varphi(n-1) = \varphi(n,\beta)$.

In the equilibria found here the thresholds are symmetric (satisfying $C_B = 1 - C_A$). If the "informative vote lead" required to reach either threshold from neutral priors ($m = 0$) is g, then this will be referred to as a g-step bandwagon. From the statement of the strategies the value of g is found by the following rule: from neutral priors g is the unique integer such that,

$$\begin{aligned} \pi(g-1) &\in (1 - C_A, C_A)\,, \text{ and} \\ \pi(g) &\notin (1 - C_A, C_A) \end{aligned}$$

An additional feature of this type of bandwagon is that for non-neutral priors ($m \neq 0$) the vote lead required by each candidate will be unbalanced and reflective of the value of m. These unbalanced bandwagons are able to describe accurately the common primary processes which start with a dominant candidate. Good recent examples of this would be the two nomination races of 2000. In both cases the front-runners, Gore and Bush, were thought to need only a few solid performances to wrap up the nomination whereas the challengers, Bradley and McCain, would only secure their place in the race with such performances.[24]

In the analysis to follow I will consider the generalized cut point voting strategy. I will look at both internal values, or BWV, and the limit cases of "uninformative" and "informative voting." These possibilities cover a broad range of possible strategic considerations that have been conjectured to occur in real sequential elections. Given the size of the voters' strategy spaces there may, of course, be other equilibrium strategies to this game. However, noting the consistency of the BWV equilibrium with observation, it is hoped that some understanding of the strategic aspects of voting in sequential elections can be gained by analyzing this general voting strategy.

[24]Judgements like this were commonplace before the New Hampshire primary. One example is from the Los Angeles Times (Brownstein (1/26/2000, p. A8)), "Conversely, if the two front-runners can reinforce their solid Iowa victories with wins in New Hampshire, the challengers may find the curtain falling in the first act.

"I think that New Hampshire is the last chance that both McCain and Bradley have to gain a toehold on the nomination," says political scientist William G. Mayer. "Even if they win, I still think the odds are probably against them. But the odds are even more strikingly against them if they lose New Hampshire.""

3.3 Results

The main result will now be stated and proven. The following theorem shows that even with "sensible" beliefs, BWV is an equilibrium for all nonzero values of k. Further, by the properties of BWV the outcomes of these equilibria require, with probability one, the start of a bandwagon. I find also that as k approaches zero the bandwagon takes longer and longer to begin. This requires a candidate to achieve an increasingly greater vote lead before all voters will support him regardless of their private signal. However, the crucial finding is that no matter how small is k a bandwagon will eventually begin.

Theorem 3.1 *Suppose that beliefs are "sensible" and that $k > 0$. Then $\exists$ $(C_B, C_A) \in (0,1)^2$, where $C_B = 1 - C_A$, such that Bandwagon Voting by all voters constitutes a perfect Bayesian equilibrium.*

Sketch of Proof: The complete proof is in the appendix. Here I will supply a sketch of the proof's structure.

1. Assume that $\pi = \frac{1}{2}$ (i.e., $m = 0$). The proof for all $m \neq 0$ follows immediately. This is a consequence of the assumption of an infinite number of voters and is a key characteristic to making the model tractable. With an infinite number of voters to follow it doesn't matter to each voter how many votes have been cast, all that matters is the respective vote counts. Thus the analysis for $m = 1$ is identical to that for $m = 0$ if we simply think of there having been one vote for candidate A before the first voter casts her ballot.

2. Similarly, I do not need to consider the incentives facing each voter in the sequence, but only a representative voter for each value of n. Symmetry implies that this can be further reduced to only α observers for $n = 0$ and all α and β observers for $n > 0$.

3. I then assume a certain length of the bandwagon, g, and show how the constraints on k for α observers are increasing in n (for $n < g$). And likewise for β

observers they are decreasing. Therefore, to satisfy incentive compatibility (IC) only the constraints on the final observers need be considered. These constraints will be denoted by $\underline{k}_{*g}$ and $\overline{k}_g$, respectively.

4. The IC constraints for voters after the bandwagon has started ($n = g$) are then determined. The constraint for β observers will be denoted by $\underline{k}_g$. The constraint on α observers is shown to be dominated by $\underline{k}_{*g}$.

5. By comparing these constraints I then show that for each g there exists a measurable region of k for which a g-step bandwagon is an equilibrium.

6. I then compare these regions for different values of g and show that for consecutive values of g they intersect. This result is then shown to imply that every value of $k > 0$ is in a region that corresponds to some g-step bandwagon, and thus the theorem is proven.

In a bandwagon equilibrium all voters vote informatively until beliefs in favor of a certain candidate pass through a threshold. After this point is reached, all subsequent voters "jump onto the bandwagon" of that candidate and vote for him regardless of their private signals. This threshold is the same for either candidate, but because of the possibility of non-neutral priors the vote lead required by either candidate to reach this threshold may differ. It is possible for the bandwagon to begin immediately, before even the first vote is cast. This will occur for sufficiently skewed priors ($|m| \geq g$). This situation could correspond to an incumbent president running unchallenged in the nomination process because it is not worth the effort for a challenger to even participate.

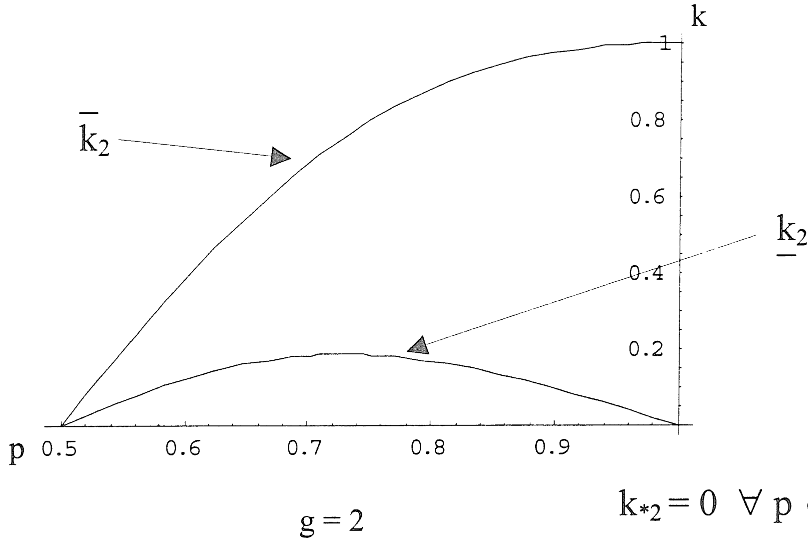
The following figures give some indication as to the length of the bandwagon that could be expected. As can be seen, the corresponding bandwagon length increases slowly as k decreases. For example, Figure 3.1c shows that for a 6-step bandwagon k must be at least less than 0.1, and even far smaller for most values of p, and Figure 3.1d implies that only for extremely small values of k is a bandwagon of length ten or more required to support an equilibrium. For each g value the suitable region of k is that between the highest curve, $\overline{k}_g$, and the second highest curve, either $\underline{k}_g$ or $\underline{k}_{*g}$. The reader should note the difference in scales between the graphs.

In the proof it is established that $\overline{k}_g$ is strictly decreasing in g. This implies that pairs of k and p above the top curve, $\overline{k}_g$ must be associated with bandwagons of shorter length than that represented by the curve. However, it is not necessarily the case that max $[\underline{k}_g, \underline{k}_{*g}]$ is decreasing in g, so the opposite statement cannot be made for points in the figure below these curves.[25] What can be said, though, is that the minimum g that is supportable as an equilibrium approaches infinity as $k \to 0$. This follows from the combination of three facts: $\underline{k}_g > 0$ for all finite g (this can be seen from the statement of $\underline{k}_g$ and is proven generally in Corollary 3.1), $\overline{k}_g > \underline{k}_g$, and $\lim_{g\to\infty} \overline{k}_g = 0$.

For any particular g the curve $\underline{k}_{*g}$ is derived from the incentive compatibility constraint for voters who observe a signal in favor of the leading candidate before a bandwagon begins. As mentioned in step 3 of the proof's sketch, the binding constraint, that represented by $\underline{k}_{*g}$, is from the final observer of a favorable signal before the bandwagon begins (i.e., when $n = g - 1$). The reader may wonder why this bound is on the upper end of possible values of n rather than the lower end, and even why these bounds are positive at all. This doubt arises because these voters are expected to vote their signal in support of the leading candidate. Therefore, intuition may suggest that they do not need the additional payoff from voting for the winner in order to be enticed into following this strategy. Further, because the belief that A is the true state and the probability of A winning the election are increasing in n, it may be concluded that if the first α observer is happy to vote for A then so too must all subsequent α observers.

However, this intuition is misleading. Strangely enough, if a bandwagon equilibrium is being played then as n increases the pure candidate selection incentive to vote for A (the non-k terms) actually decreases and for $n > 1$ it is negative. The (correct) intuition is that by voting for A and providing A with an even larger lead the voter is

[25]The non-monotonicity in g is due to the non-monotonicity of $\underline{k}_{*g}$. The intuition for this characteristic will be described briefly on page 120.



$k_{*2} = 0 \ \forall \ p \in (1/2,1)$

Figure 3.1a

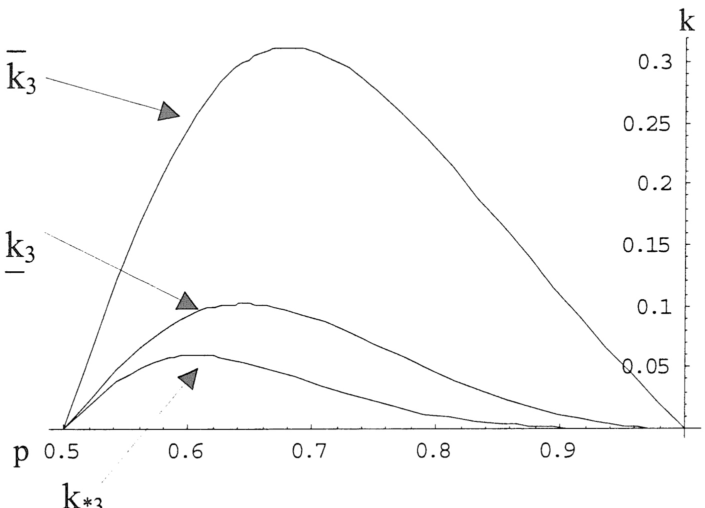
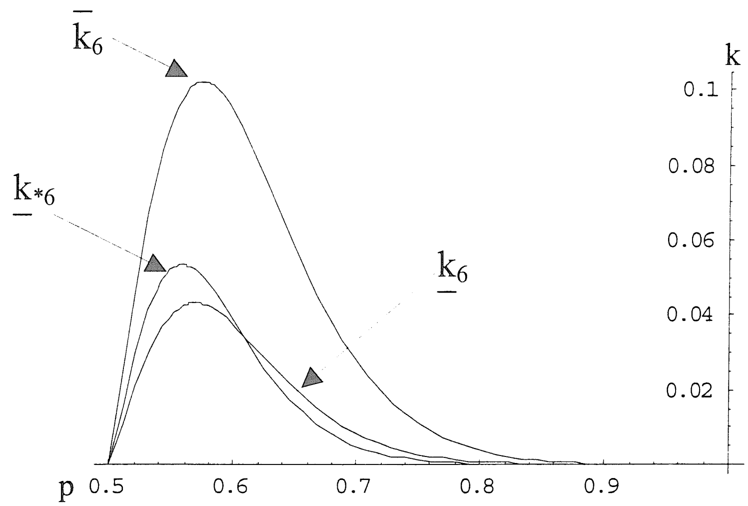


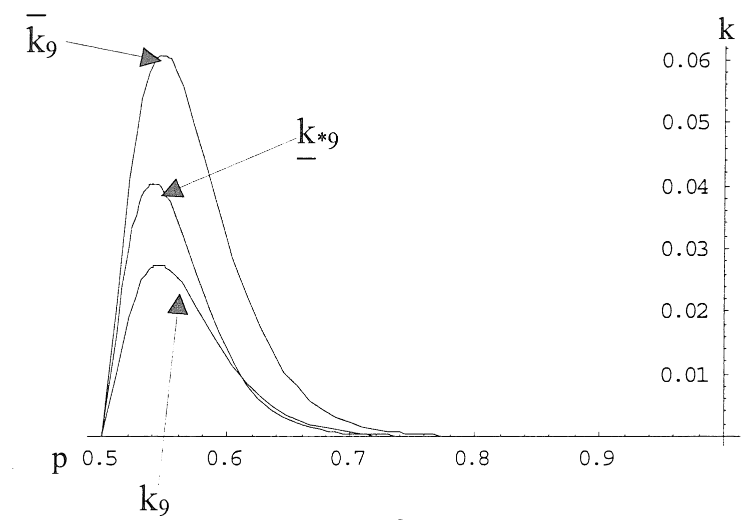
Figure 3.1b

Figure 3.1: Equilibrium Bandwagon Lengths



g = 6

Figure 3.1c



g = 9

Figure 3.1d

pushing the election closer to the bandwagon point, after which no more information is aggregated, and this proves costly. Consider an α observer when $n = g - 1$. If she votes for A, as her signal indicates, then a bandwagon begins and A is a certain winner. However, if she goes against her signal and votes for B then $n = g - 2$ and subsequent voters will continue to vote informatively. If A is the "better" candidate, this deviation will reduce only slightly the probability that A will win. But if B is the "better" candidate then his chances of overcoming the vote deficit and winning are much greater. Therefore, by deviating the voter provides more time for the "better" candidate to emerge and increases the probability that this candidate is selected (even after these probabilities are weighted by current beliefs).

Each voter trades off this informational incentive against the positive incentive to vote for the likely winner by voting with her signal. As n increases the incentive to 'go with the winner' also increases. However, the informational incentive increases at a faster rate, resulting in the tightest constraint being that of the final α observer.

This intuition does not arise in the model of Fey (1998) as he considers informational cascades which must be a length of either one or two (i.e., $g = 1, 2$). For these lengths, and only for these lengths, is it the case that $\underline{k}_{*g} \not> 0$. This finding also implies that the generalized bandwagon equilibria found here do not depend on the assumption of "sensible" beliefs. The fact that $\underline{k}_{*g} > 0 \; \forall g > 2$ implies that even without "sensible" beliefs a bandwagon of length longer than two cannot be supported as an equilibrium unless $k > 0$.

Significantly, these strange incentives are often observed in presidential primaries. An example from the 2000 campaign is provided by a comment in the Los Angeles Times (1/26/00, p. A8), "Many analysts – including some in the Gore camp – believe Bradley could benefit from a reluctance among some New Hampshire voters to possibly end the race by giving the vice president a victory." These incentives are commonly referred to, at least in the popular literature, as "buyers' remorse."

This characteristic of $\underline{k}_{*g}$ also indicates why other approaches to generalizing the voting model are unlikely to lead to the onset of bandwagons. Unless a voter has an independent incentive to vote for the leading candidate then she will vote against the leader for informational purposes. The onset of a bandwagon requires the alignment of preferences behind the leading candidate, preferences that arise, perhaps uniquely, when voters have a desire to vote for the winner.

Significantly, this implies that bandwagons can't be driven solely by the candidates or the media. An alternative conjecture to describe bandwagons may be that it is the candidates or the media, and not the voters, who decide a critical point has been reached and cease competing or covering the contest, respectively, effectively starting a bandwagon for the remaining candidate. The presence of "buyers' remorse" implies that voters would never let such a critical point be reached if they didn't have an additional incentive to vote for the winner. Thus, the incentives of the candidates or the media alone can't explain the onset of bandwagons.

Another interesting feature of $\underline{k}_{*g}$ is that it is not always monotonically decreasing in g. Significantly, as pointed out on page 116, this rules out the strong statement that any k (for a given p) that is too small for a g-step bandwagon can only support equilibria with longer bandwagon lengths. To understand why this is the case requires a closer look at "buyers' remorse" and the utility from conforming, and not conforming, with the majority decision. Suppose a g-step bandwagon is being played. If $g = 2$ then "buyers' remorse" is zero (see the proof of Theorem 3.1). This is because, with such a short bandwagon length, very little information is aggregated. Consequently, inserting some misinformation into the system (by voting against a private signal) involves a cost that does not exceed the benefit of allowing more information to be aggregated. For $g > 2$, however, this benefit exceeds the cost and the "buyers' remorse" incentive is positive. Though as $g \to \infty$ this benefit becomes small as the probability of affecting the outcome (by voting against a signal) becomes small. Thus, the strength of the "buyers' remorse" incentive is not monotonic in g. Another way of looking at this is that as g increases, the accuracy of the voting decision increases and, therefore, the ability to make informational gains gets smaller. This interpretation suggests that the "buyers' remorse" incentive will be stronger the smaller is p (as then the less accurate is the election decision for a fixed g).

		g	
		2	4
	0.6	3.846	13.402
p	0.75	3.2	7.805
	0.9	2.439	4.998

Table 3.1: Expected Number of Votes till Bandwagon Begins

In contrast, the cost of succumbing to "buyers' remorse" and voting against a private signal increases monotonically in g. This, quite simply, follows from the fact that for larger g the trailing candidate is less likely to overcome the deficit and triumph. Once again the strength of this effect depends on p. Essentially, the lower is p the greater the chance the "better" candidate loses and a vote against the strongly favored leader doesn't prove costly.

Combining these intuitions, it follows that the lower is p the stronger is the "buyers remorse" incentive and the lower is the cost of voting against the leader (in terms of k). As a result, for small values of p, the non-concavity of the "buyers' remorse" incentive dominates and persists for small integer values of g, and therefore $\underline{k}_{*g}$ is non-monotonically decreasing in g. [26]

In addition to the length of the bandwagon, it is also of interest to know the expected number of votes before the bandwagon begins. Unfortunately, a general expression for arbitrary length g does not seem tractable. As an indication, Table 3.1 presents the expected number of votes before bandwagons of length 2 and 4 begin for several possible p values.

There are several things to note from this table. Firstly, even though there are an infinite number of voters the expected number of votes until the election is effectively decided (a bandwagon begins) is small and most definitely finite. For example, even when $p = \frac{3}{4}$ and a four vote lead by either candidate is required a bandwagon is expected to begin before eight votes have been cast. Secondly, the expected number of votes is, at least for the values in the table, decreasing in p, the accuracy of the private signal. This makes intuitive sense as the more accurate the information that

[26]Numerical estimations serve to strongly support these intuitions.

is available to the voters, the more likely the better candidate will emerge, and the sooner the election will be decided.

Once a bandwagon begins, subsequent voters are required to ignore their private information and vote for the leading candidate. This entails a loss of valuable information with regards to which is the better candidate, and may lead to the least preferred candidate being selected. This result may be considered surprising in several respects. From one perspective, it is surprising that information is lost as the game is one of common interest, even when $k > 0$, as ideally all voters would like to make the same correct choice. From the opposite perspective, it may be surprising that any information is aggregated at all. Considering that after the bandwagon begins all voters (which, as was indicated in Table 3.1, is expected to be all but a finite few) will ignore their private information and vote purely in order to support the winner, it is surprising that any voters are willing to vote informatively.

The explanation of these dual intuitions is that the sequential voting mechanism creates a tension between information revelation and the desire of voters to vote for the winner, and that this tension manifests itself in an interdependent sequence of behavior. Later voters bandwagon because early voters vote informatively. This is optimal because once the bandwagon threshold is reached their incentive to vote for the winner dominates their desire to reveal their private information. Alternatively, early voters are prepared to vote informatively precisely because later voters will bandwagon. This effective correlation of votes ensures the pivot probabilities of early voters are strictly positive, even with an infinite population, and provides them with the incentive to vote informatively. Thus, whilst the sequential voting mechanism inhibits full information aggregation when voters have a desire to vote for the winner, it simultaneously ensures that not all information is lost.[27]

The potential cost of bandwagons can be quantified by determining the probability of selecting the best candidate given a bandwagon equilibrium. For a g-step

[27] That the pivot probabilities are strictly positive also indicates that all voters in an infinite population will turn out, even in the presence of a positive (but sufficiently small) cost of voting.

		g				
		1	3	6	9	∞
	0.6	.6	.7714	.9193	.9746	1
p	0.75	.75	.9643	.9986	.9999	1
	0.9	.9	.9986	≈ 1	≈ 1	1

Table 3.2: $P(W = \text{"better" candidate}|g, \pi(m))$

bandwagon equilibrium this probability can be expressed as follows.[28]

$$\begin{aligned} P(W = \text{"better" candidate}|g, \pi(m)) &= P(A|A) + P(B|B) \\ &= w(0)^{29} \\ &= \pi(g) \end{aligned}$$

Surprisingly, this expression is independent of the prior beliefs, m. Intuition can be gained by considering the conditions for a bandwagon to begin, and thus a winner to be anointed. For a bandwagon to begin either candidate must gain an "informative vote lead" such that $m + n = g$. Thus, whenever a bandwagon starts the belief that the chosen candidate is the "better" one is $\pi(g)$. And so, ex-ante, the expectation the best candidate will be picked is $\pi(g)$. Consequently, the longer is the bandwagon the more likely it is that the "better" candidate will be selected (as $\pi(g)$ is increasing in g). This probability is also equal to $w(0)$, which at $m + n = 0$ is equal to both the probability that A wins if it is the "better" candidate and the probability that B wins if it is "better." The expectation of selecting the better candidate is, therefore, $w(0)$.

The previous identities show that for any finite g there exists a chance that the inferior candidate wins the election. For low values of g this possibility can prove very costly, as displayed for some representative values in Table 3.2. Note that when $g = \infty$ a bandwagon never starts and the best candidate is almost always chosen.

[28]See Chapter 4 for a full derivation of these identities.

[29]Where $w(n)$ is derived in the proof of Theorem 3.1 and is defined as the probability that a candidate will win from neutral priors if it is the "better" candidate and has an n "informative vote lead."

The voters' desire to vote for the winner imposes a deadweight loss on themselves and society by permitting the possibility that an inferior candidate is selected. It is easy to prove from the above identities that for a particular g the probability of the inferior candidate being selected is decreasing in p. However, because g is a function of p, this does not imply that in equilibrium a larger p necessarily increases this probability. That is, an increase in p may lead to a decrease in the equilibrium bandwagon length and actually result in an efficiency loss. Critically, however, Theorem 3.2 and Corollary 3.1 will show that regardless of the value of p, $k > 0$ implies that g must be finite, and $k = 0$ implies that g must be infinite. Therefore, the voters' desire to vote for the winner results in strict inefficiency (in the sense that the "better" candidate wins with a lower probability) in the selection of candidates compared to the "informative voting" equilibrium when $k = 0$.

If the sequential election is a primary then to the members of the party a poor selection may not be entirely a deadweight loss. For example, a party may prefer a quick nomination process as provided by a short bandwagon length, even if this has the potential of leading to the selection of a sub-optimal candidate. This possibility will be discussed in Section 3.4.1. Regardless of the private motivations of the party, however, the failure to select the best candidate still imposes the cost of suboptimal government on society.

As is evident from the proof of Theorem 3.1, it is possible for certain parameter values (p and k) to be supportable as equilibria for bandwagons of different length. In this light it is important to know, assuming the CPV strategy is being played in equilibrium, whether for $k > 0$ a bandwagon will eventually commence. The next result shows that this is indeed the case. That for $k > 0$ "informative voting" (CPV(0,1)) is not supportable as an equilibrium. Thus, if a symmetric CPV equilibrium is being played it must be a BWV equilibrium, and with probability one a bandwagon must commence after a finite number of votes. Note that the following result does not require the restriction to permissible beliefs.[30]

[30] In fact, the restriction to "sensible" beliefs is also not required as every vote history can be reached with positive probability when all voters are using the "informative voting" strategy

Theorem 3.2 *Suppose that $k > 0$. Then "informative voting" (i.e., CPV(0,1)) by all voters is not a perfect Bayesian equilibrium.*

Proof: Consider a voter who faces a t vote lead for A and who observes a private signal for B. Beliefs are given by $\varphi(t-1)$, and the law of large numbers gives the following probability of victory functions.

$$w(t+1) = w(t-1) = \left\{ \begin{array}{c} 1 \text{ if A is the true state} \\ 0 \text{ otherwise} \end{array} \right\}$$

The key observation here is that because there are an infinite number of voters the probability that a voter's choice will make a difference to the outcome is zero. This gives the utility values,

$$\begin{aligned} u(v &= A|\beta, t) = \varphi(t-1)\,[1+k] + [1-\varphi(t-1)]\,[1+0] \\ &= 1 + k.\varphi(t-1) \\ u(v &= B|\beta, t) = \varphi(t-1)\,[1+0] + [1-\varphi(t-1)]\,[1+k] \\ &= 1 + k\,[1-\varphi(t-1)] \end{aligned}$$

Incentive compatibility for $k > 0$ then requires,

$$\begin{aligned} u(v &= B|\beta, t) \geq u(v = A|\beta, t) \\ &\Rightarrow \varphi(t-1) \leq \frac{1}{2} \end{aligned}$$

Which is not true for large enough t given any $\pi > 0$. If $\pi = 0$ then the same analysis for an α observer would lead to an analogous violation of IC. Thus, informative voting by all voters cannot be an equilibrium. ∎

This result shows that, no matter how small k is, if all other voters are voting informatively then a point will be reached where the chance of securing k for a voter by voting for the likely winner outweighs the information contribution of voting her signal. And thus she will "jump on the bandwagon." The crucial insight of this result is that if all voters are expected to vote informatively then in a large population the probability that any one voter will make a difference, the pivot probability, goes to zero. This is because subsequent voters cast their ballots solely on the contents of their private signal and can't be influenced by earlier voters. Consequently, the incentive to make a difference is dominated by the desire to vote for the winner.[31]

In contrast, the use of bandwagon strategies when g is finite, which also involves some voters voting informatively, is supported as an equilibrium for the very reason that not all voters vote informatively. In a bandwagon the decisions of later voters can be influenced by the votes of earlier voters. Thus, even with a large population the pivot probabilities do not go to zero. By abandoning some information (those voters who bandwagon) the use of bandwagon strategies, perhaps counterintuitively, ensures some information aggregation by allowing those voters who use their private information to make a difference. This logic proves central to the results of Chapter 4 that compares the information aggregation properties of sequential and simultaneous voting.

For a fixed p the length of the bandwagon increases without bound as k approaches zero. The following result shows that in the limit at $k = 0$ the length of the bandwagon must be infinite. That is, neither candidate can ever establish a lead big enough to start a bandwagon and so all voters vote informatively. Thus, at $k = 0$ "informative voting" is an equilibrium but BWV is not. This is the infinite voter version of the special case result of Fey (1998) and Wit (1997), and also doesn't require the restriction of permissible beliefs.

Corollary 3.1 *Suppose that beliefs are "sensible" and that $k = 0$. Then Cut Point Voting by all voters constitutes a perfect Bayesian equilibrium if, and only if, $C_B = 0$ and $C_A = 1$. Therefore, bandwagons never occur in equilibrium.*

Proof: The proof of the symmetric case, when $C_B = 1 - C_A$, and when prior beliefs are permissible will be presented here. This proof can be seen easily from the arguments

[31]For any $k > 0$ this is also true for large enough finite populations. Eventually the population will be so large that the pivot probability of a single voter is outweighed by her desire to vote for the winner. Thus "informative voting" by all voters is not an equilibrium.

in Theorem 3.1 and requires the same intuition as the general case, which has been relegated to the appendix.

{⇒sufficiency} From the proof of Theorem 3.2 the incentive compatibility constraint is the following,

$$1 + k\left[1 - \varphi(t-1)\right] \geq 1 + k.\varphi(t-1)$$

Which is satisfied by indifference if $k = 0$. Thus, all voters weakly prefer to reveal their private information and vote informatively.

{⇐necessity} Recall that $C_A \neq 1$ implies a bandwagon must begin after a g vote lead has been established by a candidate (adjusting for priors if necessary). For any g-step bandwagon, where g is finite, the values of k which support this as an equilibrium are all k such that $k \in \left[\max\left\{\underline{k}_g, \underline{k}_{*g}\right\}, \overline{k}_g\right]$. It can be seen easily from the statements of $\underline{k}_g$ and $\underline{k}_{*g}$ (in the proof of Theorem 3.1) that $\underline{k}_g > 0$ for all positive integers g, and that $\underline{k}_g$ and $\underline{k}_{*g} \to 0^+$ as $g \to \infty$. Therefore $0 \notin \left[\max\left\{\underline{k}_*, \underline{k}_{*g}\right\}, \overline{k}_g\right]$ for all integers g, and there is no g such that $k = 0$ can support a g-step bandwagon equilibrium. Thus, CPV can't be an equilibrium if $C_A \neq 1$. ∎

This result shows that the existence of bandwagons is not simply due to the assumption of an infinite voting population. That is, the continual addition of voters to the population doesn't imply that a bandwagon will eventually emerge. Rather, in conjunction with the previous results, Corollary 3.1 shows that bandwagons exist because of a combination of voters' desire to vote for the winner and an infinite population. And, further, that the nature of this bandwagon (its length) depends upon the intensity of the desire to support the winner.

Robustness

For transparency and tractability the model employs several somewhat restrictive assumptions. Two of the most notable of these are the restriction to "permissible" beliefs and that all voters enjoy a positive and identical utility from conforming to the majority decision. The robustness of the results just presented to these two assumptions will now be explored. In the conclusion I will also briefly discuss the possible impact of relaxing several other assumptions.

For the statement of Theorem 3.1 it is assumed that prior beliefs are "permissible" and given by Equation 3.1, with m required to be an integer. This is somewhat restrictive. It is also of interest to know what happens when priors take on values not representable by an integer m. Unfortunately, to solve this model for arbitrary priors becomes overly complicated as the symmetry used in the proof would be lost. However, there would seem to be no kink in the intuition of Theorem 3.1 that would indicate that the result wouldn't generalize.

But it needn't be left there. What can be seen immediately from the theorem as it stands is that the result is robust to at least small perturbations of beliefs. That is, if beliefs vary slightly from those allowable by Equation 3.1 then the theorem still holds. This observation cannot indicate whether the theorem holds for all possible priors, but it can confirm that the theorem is not an artifact of the particular set of priors that are permitted. I will have to leave as a conjecture the assertion that the theorem holds for all possible priors, but what is known is that the theorem isn't a knife edge result.

Corollary 3.2 *For any* $\varepsilon > 0$ *define* $\pi^{\varepsilon}(m) = [m - \varepsilon, m + \varepsilon]$ *where* m *is an integer. Consider any* $\tilde{m}$ *s.t. for some* m *and* ε, $\tilde{m} \in \pi^{\varepsilon}(m)$. *Now suppose that priors are determined by substituting* $\tilde{m}$ *into Equation 3.1. If beliefs are "sensible" there exists an* $\varepsilon > 0$ *such that BWV by all voters constitutes a perfect Bayesian equilibrium for all* $k > 0$.

Proof: For any g in the symmetric case the values of k that support a cascade equilibrium are given by all k such that $k \in \left[\max\{\underline{k}_g, \underline{k}_{*g}\}, \overline{k}_g\right]$. From the expressions for these values it can be seen that they vary continuously in beliefs. Therefore, as priors change the beliefs after any given voting history also vary continuously. Thus, if the priors vary continuously then the boundaries on acceptable k values for each g also vary continuously. Recalling that $\overline{k}_{(g+1)} > \underline{k}_g, \underline{k}_{g*}$, for small enough perturbations of the priors this inequality is still true. And as for perturbations it is still true that $\overline{k}_g \to 0$ the theorem must still be true. ∎

Of course, the size of the permissible interval may vary for different integer values.

The second restrictive assumption to be considered here is the homogeneity of the voters' desire to conform. To maintain uniformity (and simplicity) I have until this point assumed that all voters receive an identical benefit, $k > 0$, from conforming to the majority decision. For a variety of reasons this may not be expected to be the case in real voting situations.[32] Unfortunately, a complete relaxation of this assumption leads to intractability. However, I conjecture that such variation does not substantively alter the results generated here. To provide some intuition for why this might be the case I present here a very simple relaxation: for a finite number of voters $k = 0$, and the identity of these voters is publicly known.[33] Denote the set of such voters by R. "Buyers' remorse" implies that in this environment BWV by all voters can't be an equilibrium. Voters in R do not receive the benefit from following a BWV strategy (conforming) and instead will only feel, so to speak, the "remorse." Consequently, their tendency will be to vote against any leading candidate. I will refer to such a strategy as an "Anti-Bandwagon" strategy.

Definition 3.6 *Anti-Bandwagon Voting (ABV) is a strategy, σ_i, such that*

$$v_i = \left\{ \begin{array}{l} A \textit{ if } m+n<0 \\ B \textit{ if } m+n>0 \\ s_i \textit{ if } m+n=0 \end{array} \right\}$$

Suppose then that this strategy is used by all $k = 0$ type voters.[34] As all other voters will be aware of this their beliefs will not be affected by such votes. As before,

[32] See Section 3.4.1 for motivations of k. From these motivations it is easy to see how the level of this desire may vary across individuals.

[33] The following analysis is easily extendable to an infinite number of $k = 0$ type voters as long as their share of the population converges to strictly less than half.

[34] Given that $\underline{k}_*(1|g) = 0$ (see the proof of Theorem 3.1), the following analysis would also hold if the required levels for voting A and B were $m + n > 1$ and $m + n < -1$, respectively.

beliefs after deviations from this strategy will be "sensible." This leads to the following corollary. The proof follows simply from Theorem 3.1 and has been relegated to the Appendix.

Corollary 3.3 *Suppose that beliefs are "sensible" and that $k > 0$ for $\forall i \notin R$. Then $\exists\, (C_B, C_A) \in (0,1)^2$, where $C_B = 1 - C_A$, such that BWV(C_B, C_A) $\forall i \notin R$ and ABV $\forall i \in R$ constitutes a perfect Bayesian equilibrium.*

In this equilibrium $k > 0$ types behave as before (and the equilibrium bandwagon thresholds will be the same as for Theorem 3.1). The $k = 0$ types vote informatively if the election is tied (more precisely, if beliefs are neutral), otherwise they vote against the leader. Consequently, their actions are ignored by all voters other than when the election is tied. This leads to the rather strange conclusion that the most socially minded voters (as they care only about selecting the "better" candidate) are ignored by the rest of the population. Despite not affecting the outcome in equilibrium, these voters do not have the incentive to deviate as by doing so they can only drive beliefs further away from the interval in which informative voting takes place (and they have no incentive to conform to the majority decision to induce them to do this).[35] If there is a cost to vote then, as these voters receive no additional payoff from conforming to the majority, it may be expected that they would, if given the choice, abstain from voting.

This simple case shows that the substantive findings of the symmetric case are robust to at least some heterogeneity of the preference of agents to conform to the majority decision. Of course, with greater heterogeneity the analysis would not be nearly so simple. However, there exists no clear reason why the same intuitions would not carry over, at least in approximate form.[36]

[35] It should be noted that though this strategy requires, say in the case in which a bandwagon for A has begun, α observers to vote for B, this is not a weakly dominated strategy.

[36] If we expand the model only slightly then the complication increases significantly. Suppose that in addition to the two types of the corollary there were also voters with small but positive k. If k is too small then in equilibrium these voters can't be following the BWV strategy (that the large k voters are following). However, neither can they be playing the ABV strategy in equilibrium. If playing the ABV strategy then these voters would wish to deviate and support the leading candidate

Similar conclusions are drawn if a voter's desire to conform is instead private information. In this case a vote against a leading candidate would be discounted for the possibility that the vote was from a $k = 0$ type. Therefore, the presence of the $k = 0$ types affects the utility calculations of the $k > 0$ type voters and leads to excessive complication. However, it would seem that, despite this complication, $k > 0$ type voters would still be inclined, if the lead became substantial, to abandon their private information and support a leading candidate.

Related Literature

In addition to Fey and Wit, Dekel and Piccione (2000) also consider the possibility of cascades in sequential voting. Like the other papers, they assume a finite population and consider only the limit case of $k = 0$. They point out that when priors are non-neutral an equilibrium exists in which some voters vote uninformatively for the favored candidate and all others vote informatively. They claim that this equilibrium displays, "seemingly cascade behavior" (p. 36) because all of the uninformative voting must come at the end, and it may begin before either candidate establishes a majority. Observationally the play of this equilibrium may at times look like a voting cascade, though as the candidate who benefits from the uninformative voting is not determined by previous voting it is not in the spirit of voting bandwagons described in the empirical literature. Additionally, these cascades exhibit several characteristics which would suggest that they are fundamentally different to the bandwagons observed in sequential elections.

The primary discrepancy is that a Dekel and Piccione cascade can only start for the candidate who begins the election as the front runner. This implies that the "underdog" candidate can never perform well enough to benefit from a cascade. There are many counterexamples to this characteristic, most notably Carter in the Democratic primaries of 1976.[37] This inability is not only a trait of "underdogs," but

as, given the current strategies, they have no informational impact voting against him. In this case "Sensible" beliefs requires a deviation against the leader to then be informative, which induces these voters to switch again and vote against the leader. Thus, any equilibrium in this environment would be significantly more complicated and require some mixing for these low k types (β observers always vote B, and α observers mix over A and B).

[37]See, amongst others, Aldrich (1980).

of all non-front runners. Therefore, if priors are neutral a cascade will never be seen. In fact, if the population becomes large then the priors must become increasingly skewed for the probability of a cascade occurring to not approach zero. Thus, the characteristics of the cascade type behavior described by Dekel and Piccione do not correlate fully with the frequency and nature of bandwagons observed in sequential elections.

3.4 Discussion

3.4.1 Motivation: Why Vote for the Winner?

There are many possible reasons why a voter would place utility on whether the option or candidate she supports is likely to win the vote. These motivations can be of a personal, psychological foundation or they can be rational calculations once the full context of the voting game is taken into account. As I prove my results for all nonzero weightings on this payoff it doesn't matter whether all of these motivations are present, or even whether they are strong or weak. All that is important is that they affect the voting decision to some degree. In that light, and considering the pervasiveness of the motivations it would seem that this extension of the utility function is a natural and valid one.

Of course the existence of these motivations doesn't necessarily imply that they are actually employed in the vote decision. However, the validity of the extension is confirmed by Bartels (1988) in his study of U.S. presidential primaries when he concludes,[38]

> "The overall impact of expectations seems to involve at least three distinct effects reflecting distinct psychological processes in the minds of prospective voters - processes ranging from rational strategic calculation

[38]More recent evidence consistent with a desire to vote for the winning candidate is provided by Herron (1998). In a study of the 1992 U.S. presidential election he finds that Clinton supporters were significantly more likely to turn out and vote if they thought he was going to win, and Bush supporters were significantly less likely to turn out if they held the same beliefs. Effectively, Clinton supporters were inspired to vote for a winner and Bush supporters were inspired to abandon a loser.

to a simple, uncritical desire to "go with a winner." In combination, these distinct individual-level effects make up the very significant aggregate level effect of expectations on choices that campaign observers call "momentum."" (p. 10)

Rational Choice Motivations

Often the sequential voting game specified here is part of a larger strategic environment, and when the complete game is considered a desire to vote for the winner on the part of voters may arise as a traditionally rational, utility maximizing choice. There are many possible extensions that would be capable of producing this result. Two extensions that are very plausible, and that require little additional consideration will be presented here.

The most obvious possibility arises from the fact that the winner of the election typically holds some degree of arbitrary power over the voters. If the winner abuses this power to reward his supporters (indeed, it may very well be in his interests to do so) then voters would wish to have voted for the winner, whoever it may be. There are many examples of voting environments where such incentives would arise. Politics is perhaps the most obvious as the winner controls the government. However, it may be in more personalized situations that this incentive is at its strongest. For example, workplace superiors hold the power of employment over underlings and this power may exert undue influence on behavior in decision situations and lead to conforming "yes men" type behavior.

A more subtle possibility arises if the sequential election is a precursor for a further election, as is the case with the U.S. presidential primaries. It is argued that this secondary stage impacts the objectives of voters (not to mention candidates) in the initial primary stage, and impacts them in such a way that the voters act as if they place utility on voting for the likely winner.

Aldrich (1980) claims that in a close race, even if both candidates have shown the ability to win votes in the subsequent general election, the voters are better served to discard their private information and instead bandwagon on either of the two candidates in order to give the chosen candidate, and thus the party in general, the best possible chance to win office. He describes the incentive as follows.

> "Winning a close race for the presidential nomination, on the other hand, might decrease the odds of winning the general election." (p. 14)

He goes on to argue that this implication played a crucial role on the outcome of the presidential race between Ford and Carter in 1976.

Behavioral Motivations

Casual empiricism in virtually any aspect of society would suggest that people like to win. It is likely, therefore, that these desires carry over to the political world and motivate not only candidates but voters as well. Just as a sports fan wants to support a winning team, so too does a voter want to support a winning candidate.

This behavioral motivation is further supported by the social psychology literature on conformism.[39] This literature shows that individuals have a tendency to conform to group decisions, even when they privately believe the group decision to be wrong. This interpretation is consistent with the desire to vote for the winner as in a majority rule election a vote for the winning candidate, by definition, conforms with a majority of other voters.

These motivations may also arise if voters perceive a vote for the winner to be a signal of intelligence, or some other desirable trait. Indeed, such a perception would even be appropriate if the model represents an environment in which all voters receive the same signal but only a fraction p, the intelligent fraction, perceive this signal in the correct way.

Combining these intuitions there would appear to exist both ample rational choice and behavioral reasons for why a voter would want to vote for the winning candidate. What is important here is not which forces are at work, but just that some positive incentive to 'go with a winner' exists. Given the above justifications this presumption would seem reasonable.

[39] An extensive review of the literature is contained in the text by Aronson, Wilson and Akert (1997).

3.4.2 Is this an Appropriate Model?

In addition to explaining the existence of bandwagons there are other characteristics of the primary process that are consistent with the assumptions and bandwagon equilibrium of the model.

Policy

The model presented here is without policy content. Instead the candidates compete purely on what may be considered quality attributes. This assumption is not too strange in the context of U.S. nomination campaigns and, in fact, it may even be the most appropriate assumption. It is a common empirical finding that policy matters little in primary elections. In summarizing this finding Bartels (1988, p.83) suggests that this is the case, even though policy is important to the voters, because the candidates within any one party tend to converge in their positions. Indeed, even if there are intra-party differences in policy they are likely to be small compared to the difference with the candidate to be faced in the general election. Consequently it is likely to be in each voter's interests to vote for the candidate from her party most likely, if nominated, to win the general election.

A further reason why primary voters may not be interested in policy is because it is not in their own interests to elicit firm policy positions from their candidates. Instead they would prefer their candidates to be ambiguous at the nomination phase to ensure they have full flexibility to compete at the general election. Regardless of which of these, or any other, reason is the most accurate it would seem an absence of policy positions in the description of the candidates is not an inappropriate abstraction of the actual primary process.

Voter Expectations

Another empirical regularity is that voters form different expectations as to which candidate is likely to win. The model of asymmetric information assumed here is consistent with this. Unfortunately, the binary signalling space of the model is not rich enough to account for the diversity of expectations found in real primaries.

The assumption that the signals are independent and that each is observed by a single voter is also an approximation, but one that isn't completely unrealistic. This startling fact is explained by Popkin (1994, p.119) who points out that until the week of the election in their state many primary voters are not even familiar enough with the candidates to give them a general favorability rating.[40]

Voter Behavior

When playing a BWV equilibrium the only information each voter needs in order to make her vote decision is her private signal and the vote history (in fact, all she needs is the current vote difference, n). And once the bandwagon has begun the voter needn't even bother about her private signal! This is highly consistent with observations of the information level of voters when they come to make their decision. Popkin (1994) presents evidence that primary voters do learn 'horse race' information about the progress of the nomination battle but very little else until their state's primary.

Within the strict confines of the model this apparently simplistic voting behavior is fully rational. This is because the current vote difference is a sufficient statistic for all publicly available information. However, in the real world voters could feasibly pay more attention to the campaigns and thus observe more 'private signals' than just their own. This could change voter strategies and possibly lead to better decisions by the electorate. The trade-off would be the increased expense of gathering this information. The point here is that even though U.S. primary voters typically search for little information, they are acting rationally for this information level by employing such simplistic strategies.

[40] This unusual timing of signals also abrogates the criticism that simple pre-play communication would circumvent the information aggregation problems of the model. This is a frequent criticism of common interest games. However, if each voter learns her signal only when it is her turn to vote, as is argued by Popkin, then such pre-play communication would not be possible.

Learning

Recent work by Alvarez and Glasgow (1997) suggests that voters do learn during presidential campaigns (both during the nomination phase and the general campaign). Further, they find that voters incorporate new information into their beliefs via Bayesian updating. This is exactly what voters in this model are doing when they calculate π and φ.

Given this consistency the equilibrium of the model suggests a link from voter learning to how voters make their decisions, thus providing a potential explanation for how campaigns can matter.

3.4.3 Momentum

The concept of "momentum" has taken on almost mystical tones in its application to sequential voting. It has been used to describe many different and contradictory phenomena. Often it is even used as a synonym for a bandwagon. It reached such a point that Reeves (1977) was forced to define it as, "the political cliche used to describe what is happening when no one is sure." (p.180) Researchers have made many attempts to define momentum. A loose definition, but one that encompasses the notion of many others was suggested by Aldrich (1980) when he observed, "Candidates may be said to have momentum if their chances are improving." (p.103)

Typical explanations of momentum appeal to psychological as well as informational forces. By examining the implications of the BWV equilibrium an explanation of momentum arises that is purely mechanical. That is, it explains the empirical phenomena associated with momentum purely through Bayesian updating and conditional probabilities. In a bandwagon voting equilibrium a candidate's chances of victory can be represented by a simple mathematical expression. If candidate A holds an n "informative vote lead" then his probability of victory is given by

$$P\left(A \text{ wins}|n\right) = \pi\left(n\right) w(n) + \left[1 - \pi\left(n\right)\right]\left[1 - w\left(-n\right)\right]$$

This expression is the probability that A wins when he is the "better" candidate plus the probability that A wins when B is "better." It is easy to see that a candidate's probability of victory is increasing in n, and thus when his vote lead is increasing he has momentum.

This interpretation of momentum provides a formal distinction between the concepts of momentum and bandwagons in sequential voting. It follows from this interpretation that momentum itself isn't detrimental to the quality of vote choice. Rather, it is only when this momentum turns into a bandwagon that valuable information is lost.

Despite their differences, most studies of momentum identify three defining features.[41] Firstly, after a primary victory the support for the victor typically increases in subsequent national polls. Secondly, this increased support is likely to lead to an increase in the vote in subsequent primaries for the leading candidate. Finally, and what has been the most elusive to explain, momentum can reverse itself. That is, one candidate can be gaining momentum when suddenly it reverses and flows against the candidate.[42] This implies that the gaining of momentum does not lead to inevitable victory for the beneficiary. The momentum that arises here out of the BWV equilibrium is consistent with each of these regularities. They will now be discussed in turn.

The first characteristic is easy to explain. Consider the situation where the election begins with neutral priors ($\pi = \frac{1}{2}$). By definition all voters hold this belief and so should any outside observers (in the real world this would be the media and so forth). Now consider if the first voter votes for candidate A (assuming $g > 0$). All voters infer from this that voter 1 had received a signal of α and update their beliefs to $\pi(1) = p$. Similarly, their expectation that the leading candidate will win the nomination mechanically increases from $P(A \text{ wins}|n = 0)$ to $P(A \text{ wins}|n = 1)$, as is empirically observed. Thus, if voters are asked to name which candidate they prefer, or even which they think is the most likely to win, then their answer will mechanically

[41] See Bartels (1988) for a discussion of each of these characteristics.

[42] George Bush's rapid rise and pullback early in the 1980 Republican campaign is a commonly cited example.

increase due to the outcome of the first primary.[43]

The second characteristic can't be explained precisely by the model because of the assumption that a representative voter casts a single ballot in each period, as this permits only binary vote shares. However, the model can still capture this characteristic by predicting a mechanical correlation in the performance of candidates in subsequent primaries. That is, victory in one primary increases the chances of further victories. This occurs because even though the private voter signals are independently drawn they are correlated in the sense that given a certain candidate is "better" a signal in favor of that candidate is more likely to be drawn than one favoring the other candidate. Consequently, after observing the first signal (and corresponding vote) it is more likely that the next signal will be the same and another vote for this candidate will be seen. The mathematics of this is quite simple. Given there have been two signals (and two votes), then for any prior, π,

$$\begin{aligned} P(s_1 = s_2) &= \pi\left[p.p + (1-p)(1-p)\right] + (1-\pi)\left[p.p + (1-p)(1-p)\right] \\ &= p.p + (1-p)(1-p), \text{ similarly,} \\ P(s_1 \neq s_2) &= p(1-p) + (1-p)p \end{aligned}$$

which can be rearranged to

$$\begin{aligned} P(s_1 = s_2) - P(s_1 \neq s_2) &= (2p-1)^2 > 0 \\ &\Rightarrow P(s_1 = s_2) > P(s_1 \neq s_2) \end{aligned}$$

Thus, from independent draws it is more likely that a vote for one candidate will be followed up by another vote for the same candidate rather than a reversal and a vote for the other candidate (though, of course, it is very possible).

Finally, as momentum is evolving with each informative vote it is easy to see in

[43]The expectations and preferences for a voter who hasn't yet cast a ballot also vary in this way, dependent upon her private signal. Therefore, surveys early in the campaign, before most voters have cast their ballots, should find a correlation between these two variables. Such a correlation is consistent with the empirical findings of Bartels (1985).

the model that momentum can reverse itself. Consequently, its development does not lead to inevitable victory. This is because voters vote informatively until the bandwagon threshold is reached and thus all that is required for momentum to be reversed is an unfavorable signal. However, once a bandwagon begins a victory for the leading candidate follows inevitably. It is at this point that information is no longer aggregated into the group decision. Thus, it is the onset of bandwagons and not momentum that should cause us to fear that a "mindless following of the mob"[44] is ruining the democratic process.[45]

3.5 Conclusion

I have shown that once the utility function is extended in a natural way by including a voter's desire to vote for the winning candidate the bandwagon voting strategy is supportable as a Perfect Bayesian Equilibrium. And as long as this incentive is nonzero bandwagons will occur with probability one. Further, informative voting in these instances is not equilibrium behavior.

In contrast to previous work, I employ a generalized definition of a bandwagon strategy which allows the bandwagon threshold to vary with the parameters, k and p, the utility from voting for the winner and the precision of the private signals, respectively. This implies that the number of votes a candidate needs to benefit from a bandwagon can vary from election to election, a characteristic of bandwagons consistent with empirical observation. Another empirical observation is that the required vote lead may be uneven for competing candidates. This too can be explained by the model as in a Bandwagon Voting equilibrium with non-neutral priors such unevenness is an equilibrium characteristic (though the absolute level of information is identical for the two candidates the required accumulation differs if they start from different levels).

The generalized definition of cut point voting encompasses "bandwagon voting,"

[44]Popkin (1994, p. 117).

[45]Surprisingly, in a comparison with simultaneous voting, I show in Chapter 4 that the possibility of bandwagons may actually prove to be beneficial in certain circumstances.

"informative voting" and "uninformative voting" as special cases, thus drawing together what were previously thought to be three distinct voting strategies. In equilibrium "informative voting" is the limit case in which a bandwagon never starts, and "uninformative voting" corresponds to the other extreme in which the bandwagon starts immediately. This allows us to see the results of Fey (1998) and Wit (1997) as a natural limiting case of a more generalized bandwagon model, and one in which bandwagons start at all but this limit point.

I presented an explanation of momentum which follows naturally from the play of the bandwagon equilibrium. I showed how the characteristics of momentum arise mechanically from the model of incomplete information employed here. That simply as a consequence of gradual information revelation a vote for a candidate can lead to greater voter support and a perception of improved performance in subsequent primaries. Further, the momentum described here can be reversed. This interpretation of momentum provides a formal separation and distinction between the often ambiguous concepts of momentum and bandwagons in sequential elections.

I will conclude the chapter with a series of comments on assumptions made herein and directions for future research. It may be wondered if the bandwagon equilibrium found here also exists when the voting population is finite. I conjecture that it would and see no obvious reason why the intuition would not continue to a finite population. I would suspect that the bandwagon length is not constant across time as it is with an infinite population. That is, say there are 100 voters then the 10th voter may decide to bandwagon only if candidate A has at least a seven vote lead, but the 20th voter may only require a six vote lead, the 40th a five vote lead, and so on. Obviously the calculation of these equilibria would be critically dependent on voter number, and that is why I generalized by assuming an infinite population. However, these equilibria for finite populations can be calculated and hopefully the framework and techniques introduced here can be used to determine the equilibrium for any particular application.

In an encouraging sign it can be seen immediately that if $g = 1$ the equilibrium found here also holds for any finite population. The reason for this is the simplicity of the end conditions for these particular bandwagons. With the exception of the last voter the calculations for each of these voters is unchanged whether the population is finite or infinite. Unfortunately, for larger g a finite population complicates the end conditions such that a voter's decision depends on her position in the sequence. Consequently the general solutions found here do not carry over directly. However, by simple calculations on $\underline{k}_1$ it is clear that there are many parameter values for which $g = 1$.[46] Thus, for these values a bandwagon equilibrium must exist, even when the population is finite.

Even though most, if not all, nomination campaigns are whittled down to two candidates, for large parts of the process, particularly at the beginning, there are more than two candidates competing. In this light the assumption of two candidates made here is restrictive. This criticism is valid, and hopefully future work can deal with this extension. An initial conjecture may be that the logic of the two candidate model carries over. That is, there may exist multiple thresholds that the beliefs about the trailing candidates must exceed for these candidates to receive informative votes and stay in the race. Therefore, as with the two candidate race that is eventually whittled down to one candidate, these multiple candidate races will be continually whittled down. Significantly, if true this result would be in contrast to the "folk theorem" type results of the simultaneous voting literature whereby voters focus in on any pair of candidates (see, for example, Palfrey (1989) and Fey (1997)). This possibility warrants further investigation. However, regardless of the equilibria that are found, once a campaign is whittled down to two candidates the model presented here becomes applicable. And even during the multiple candidate phase, though coordination problems and other strategic considerations may also be present, the dynamic forces captured here should still be relevant and useful in understanding the vote choices made.

It would seem unlikely that all candidates would be equally adept at conveying signals to the electorate. Similarly, it would seem unlikely that voters are equally adept at perceiving these signals. Consequently, it would be concerning if the results

[46] For $g = 1$ the constraint $\underline{k}_{*1}$ does not exist.

presented here were not robust to variations in the precision of signals, both across candidates and across voters. This question requires further investigation, though the intuition gained from the model would suggest that the results are robust to this flexibility, at least in approximate form. A reasonable conjecture would be that if the precision of signals varied across candidates then so too would the vote lead required by each candidate before a bandwagon would begin. Similarly, as precision varied across voters then so too may their behavior. A possibility would be that voters with low precision signals would bandwagon earlier than high precision voters. However, despite this asymmetry there would seem to be no obvious reason suggesting that the intuition about thresholds of support and bandwagons would not still be valid.

This type of asymmetric equilibria may also be the case if the utility from voting for the winner varied across individuals.[47] Just as the bandwagon length varies with the collective k then so too may the bandwagon thresholds of individual voters as their personal desire to vote for the winner varies. Additionally, this heterogeneity may also affect the willingness of voters to follow the BWV strategy. The fact that $\underline{k}_{*g} > 0$ implies that for $n > 1$ voters with low k always wish to vote against the leading candidate. This leads to the counterintuitive conclusion that the most 'rational' voters, those who receive no additional utility from voting for the winner, vote uninformatively and, consequently, are ignored by the other voters.

However, the more interesting extension from varying k is the possibility that individual voters not only have a desire to vote for the winner but also a fear of voting for the loser. In the current model this extension has no effect, but if voters have the option to abstain then it becomes critical. It is possible that voters will abstain because of the fear of making an unpopular choice. These motivations suggest a possible link between the level of voters' uncertainty and their decision to turnout. I hope to explore this possibility in future work.

Another reasonable variation of the model is to instead assume that voters gain less, or even zero, additional utility from voting for the winner once the race has effectively been decided. That is, k decreases once $m + n = g$ and a bandwagon

[47] This possibility was discussed briefly on page 129.

begins. An interpretation of this variation is that voters like to 'make a difference,' and therefore derive utility not from voting for the eventual winner but rather from helping him win. If this desire, k, is lowered below the threshold for the g-step bandwagon then these voters will have an incentive to deviate and vote against the bandwagon. However, despite these incentives the bandwagon equilibrium found here will still exist, at least in approximate form! As the later voters have the incentive to vote against the bandwagon regardless of their private signal when $g > 2$, such a deviation cannot be informative. Therefore, once a bandwagon begins it cannot be restarted. And due to the existence of "buyers' remorse" incentives the voters before the bandwagon begins still require $k > 0$ to follow the equilibrium strategy. Significantly, I can prove this altered version of the theorem for all possible priors, and thus the restriction to "permissible" priors is not needed.[48]

The equilibrium just described exhibits several additional features that correspond to observations from late in the presidential primary season. Firstly, as voters after the bandwagon begins will want to vote against the leader regardless of their private signal, they will be ignored by the rest of the voters. Such voters are often observed in real primaries after the nomination has effectively been secured, but they fail to reignite the race. If these voters have a positive cost to voting then given the option they will choose to abstain. This too is verified empirically as late in the nomination campaigns turnout declines considerably. Finally, this suggests that a voter's utility is greater if they vote early when they can still affect the outcome. Such a desire can also explain the effective front-loading of the U.S. presidential primary system that has been observed in recent times.

As a final note, to determine whether bandwagons actually exist in real voting populations, and to determine their characteristics, we can turn to experiments. While there has been significant experimental work looking at simultaneous voting in a jury context (see, for example, Guarnaschelli, McKelvey and Palfrey (2000)) there is little

[48]The proof of this claim employs the same techniques as used here. For tractability I assume that once a bandwagon begins k declines to zero. A similar conclusion (i.e., deviations against the bandwagon are ignored) would arise if, alternatively, the model permitted the level of k to be voter specific and private information.

work on sequential voting. Work has been done in the context of primaries by Morton and Williams (2000) and Cherry and Kroll (2000), but neither of these papers examines the possibility of bandwagons directly. In an economic context, recent work by Hung and Plott (2000) considers sequential decision making when agents have an explicit incentive to conform. They found greater evidence of cascades in this environment. Their work makes one optimistic that the utility extension implemented here will result in bandwagons in a sequential voting game.

3.6 Appendix

3.6.1 Proof that $P(\text{BW starts}|\text{BWV}) = 1$

Assume that the bandwagon is of length g. After $2g$ votes the probability a bandwagon has started is at least $p^{2g}+(1-p)^{2g}$. Thus, the probability a bandwagon hasn't started is bounded above by $1-\left[p^{2g}+(1-p)^{2g}\right]$. Similarly, as voters $2g+1$ to $4g$ cast their ballots the probability that a bandwagon begins during this period is at least as great as $p^{2g}+(1-p)^{2g}$. Therefore, the probability that a bandwagon hasn't begun after $4g$ voters is bounded above by $\left[1-\left[p^{2g}+(1-p)^{2g}\right]\right]^2$. Repeating this process implies that the probability a bandwagon hasn't started after dg voters is bounded above by $\left[1-\left[p^{2g}+(1-p)^{2g}\right]\right]^{\frac{d}{2}}$. As $p \in \left(\frac{1}{2}, 1\right)$ then $\lim_{d\to\infty}\left[1-\left[p^{2g}+(1-p)^{2g}\right]\right]^{\frac{d}{2}} = 0$. Thus, the probability that a bandwagon eventually starts is one. Note that this proof also applies if the thresholds aren't symmetric (if the thresholds require r and s votes to be reached then set $2g = r + s$ and proceed as above).

3.6.2 Proof of Theorem 3.1

Assume that $\pi = \frac{1}{2}$ (i.e., $m = 0$). The proof for all $m \neq 0$ follows immediately as there is an infinite number of voters. With an infinite number of voters to follow it doesn't matter to each voter how many votes have been cast, all that matters is the respective vote counts (as this fully describes beliefs at any point). Thus the analysis for $m = 1$ is identical to that for $m = 0$ if we simply think of there having been one vote for candidate A before the first voter casts his ballot. The only additional case that needs to be considered is when $|m| > g$. WOLOG assume that $m > g$. If a voter observes a signal of α then she faces the following two choices (recalling that even if he deviates all other voters will follow their BWV strategies).

$$\begin{aligned} u_i(v_i &= A|\alpha, 0) = \varphi(m+1) + k \\ u_i(v_i &= B|\alpha, 0) = \varphi(m+1) \end{aligned}$$

Therefore, for $k > 0$ each voter strictly prefers to follow the bandwagon. If β is observed the same expressions hold with $\varphi(m-1)$ replacing $\varphi(m+1)$ and, similarly, the voter strictly prefers to follow the bandwagon.

Voter i faces the following two choices if $m = 0$ and she has observed β (recalling the definition of n from Equation 3.2).

$$\begin{aligned} u_i\left(v_i = A|\beta, n\right) &= \varphi(n-1)[w(n+1)(1+k)] \\ &\quad +[1-\varphi(n-1)][w(-n-1) + (1-w(-n-1))k] \\ u_i(v_i = B|\beta, n) &= \varphi(n-1)[w(n-1) + (1-w(n-1))k] \\ &\quad +[1-\varphi(n-1)][w(-n+1)(1+k)] \end{aligned}$$

where $w(n)$ is the probability that A, given it is the true state, wins the election when it has an n informative vote lead over B. By the symmetry of the problem the same definition holds for when B leads A. Thus, an expression for w must be derived in order to compare the available choices and derive equilibrium conditions. It is to this purpose that I now deviate to state and prove the following lemmas.

Lemma 3.1 *For* $\forall n \in \{-g, -g+1, ..., g-1, g\}$, *where* $g \in N$,

$$w(n) = \frac{1-\left(\frac{1-p}{p}\right)^{n+g}}{1-\left(\frac{1-p}{p}\right)^{2g}}.$$

Proof: The structure of this function can perhaps be seen more clearly if the problem is abstracted from the voting model at hand. As all voters vote informatively until the bandwagon boundaries are reached the derivation of w is identical to the following problem: a gambler starts with $\$n$ and bets on a biased coin with probability of a head given by $p > \frac{1}{2}$. He bets even money, \$1 per flip, and stops when he has reached $\$g$ in his pocket or he owes $\$g$. What is the probability that he wins? Let $w(n)$ be this probability for all integers $n \in [-g, g]$. The boundary conditions are $w(-g) = 0$ and $w(g) = 1$. Suppose, for the moment, that $w(-g+1) = \delta$. By simple conditioning on cases it follows that w has the following recursive relationship.

$$w(n) = p.w(n+1) + (1-p).w(n-1) \tag{3.3}$$

This rearranges to,

$$w(n+1) = \frac{w(n) - (1-p)w(n-1)}{p}$$

which is best expressed in matrix form. Define, $W(n) = \begin{bmatrix} w(n+1) \\ w(n) \end{bmatrix}$, and $A = \begin{bmatrix} \frac{1}{p} & -\frac{1-p}{p} \\ 1 & 0 \end{bmatrix}$. Then, $W(n) = A.W(n-1)$. As $w(-g) = 0$ and $w(-g+1) = \delta$, $W(-g) = \begin{bmatrix} \delta \\ 0 \end{bmatrix}$. Combining the previous two expressions gives, $W(n) = A^{n+g} \begin{bmatrix} \delta \\ 0 \end{bmatrix}$. Powers of A are computed by diagonalization. The characteristic equation is, $\lambda^2 - \frac{1}{p}\lambda + \frac{1-p}{p} = 0$, which simplifies to, $(\lambda - 1)\left[\lambda + 1 - \frac{1}{p}\right] = 0$. Therefore, the eigenvalues are $\lambda = 1, \frac{1-p}{p}$. Simple algebra shows the corresponding eigenvectors to be [1 1] and $\left[\frac{1-p}{p} \ 1\right]$. I can now write, $D = P^{-1}AP = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1-p}{p} \end{bmatrix}$, where, $P = \begin{bmatrix} 1 & \frac{1-p}{p} \\ 1 & 1 \end{bmatrix}$ and $P^{-1} = \left(\frac{1}{1-\left(\frac{1-p}{p}\right)}\right) \begin{bmatrix} 1 & -\left(\frac{1-p}{p}\right) \\ -1 & 1 \end{bmatrix}$. Taking the power of A,

$$\begin{aligned} A^{n+g} &= PD^{n+g}P^{-1} \\ &= \left(\frac{1}{1-\left(\frac{1-p}{p}\right)}\right) \begin{bmatrix} 1 & \frac{1-p}{p} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \left(\frac{1-p}{p}\right)^{n+g} \end{bmatrix} \begin{bmatrix} 1 & -\left(\frac{1-p}{p}\right) \\ -1 & 1 \end{bmatrix} \\ &= \left(\frac{1}{1-\left(\frac{1-p}{p}\right)}\right) \begin{bmatrix} 1 - \left(\frac{1-p}{p}\right)^{n+g+1} & -\left(\frac{1-p}{p}\right) + \left(\frac{1-p}{p}\right)^{n+g+1} \\ 1 - \left(\frac{1-p}{p}\right)^{n+g} & -\left(\frac{1-p}{p}\right) + \left(\frac{1-p}{p}\right)^{n+g} \end{bmatrix} \end{aligned}$$

Recalling that $W(n) = A^{n+g} \begin{bmatrix} \delta \\ 0 \end{bmatrix}$, the second term can be extracted to find, $w(n) =$

$\left(\frac{1}{1-\left(\frac{1-p}{p}\right)}\right)\left[1-\left(\frac{1-p}{p}\right)^{n+g}\right]\delta$, which is linear in δ. Using the fact that $w(g)=1$ and solving for δ gives, $\delta=\frac{1-\left(\frac{1-p}{p}\right)}{1-\left(\frac{1-p}{p}\right)^{2g}}$. Substituting back in produces the solution,[49] $w(n)=\frac{1-\left(\frac{1-p}{p}\right)^{n+g}}{1-\left(\frac{1-p}{p}\right)^{2g}}$. ∎

From this identity the following useful relationships can be derived. These relationships will be used during the proof to simplify expressions.

Lemma 3.2 *For* $\forall n\in\{-g,-g+1,...,g-1,g\}$, *where* $g\in N^{++}$,

$$w(n+1)-w(n)=\delta\left(\frac{1-p}{p}\right)^{n+g}, \text{ and } w(n+2)-w(n)=\left(\frac{\delta}{p}\right)\left(\frac{1-p}{p}\right)^{n+g}.$$

Proof: I will derive these relationships from the identity in Lemma 3.1. They can also be derived directly from the recursive condition of Equation 3.3. Taking differences and simplifying produces, $w(n+t)-w(n)=\frac{1-\left(\frac{1-p}{p}\right)^{n+g+t}}{1-\left(\frac{1-p}{p}\right)^{2g}}-\frac{1-\left(\frac{1-p}{p}\right)^{n+g}}{1-\left(\frac{1-p}{p}\right)^{2g}}=\frac{\left(\frac{1-p}{p}\right)^{n+g}\left[1-\left(\frac{1-p}{p}\right)^{t}\right]}{1-\left(\frac{1-p}{p}\right)^{2g}}=\left(\frac{1-p}{p}\right)^{n+g}w(-g+t)$. The relationships of the lemma are then two simple variants of this general relationship. Therefore, $w(n+1)-w(n)=\delta\left(\frac{1-p}{p}\right)^{n+g}$, and, $w(n+2)-w(n)=\left(\frac{\delta}{p}\right)\left(\frac{1-p}{p}\right)^{n+g}$, as $w(-g+2)=\frac{\delta}{p}$ (from the recursive relation). ∎

For a fixed p I will now assume that a g-step bandwagon equilibrium exists and show that there is a measurable interval of k values that is consistent with this equilibrium.

With an infinite number of voters, all of whom must have incentive to vote according to their equilibrium strategies, it may be suspected that there are an infinite number of constraints to satisfy. This is true, but I will proceed by showing that for a given g these can be reduced to four conditions that need to be compared. With an infinite number of voters to follow, the decision facing each voter is independent

[49] After arriving at the final expression a simpler proof becomes apparent. Using the fact from the characteristic equation that $p\lambda=\frac{\lambda-(1-p)}{\lambda}$ and $pv=\frac{v-(1-p)}{v}$ then it is relatively easy to show that $\tilde{W}(n)=\lambda^{n}-v^{n}$ satisfies the recursive condition. This is already zero at the lower boundary, as desired. Normalizing at the upper boundary makes it into a probability.

of her position in the sequence and only depends on the "informative vote lead" and her private signal. Thus, I need not consider the incentives facing each voter but rather only those facing a representative voter for each value of n. Further, by the symmetry of the problem only the incentives of α and β observers when candidate A has a vote lead ($n > 0$), and an α observer when vote counts are tied, need to be considered. Consider β observers when A has an n vote lead in a g-step bandwagon, where $n \in \{1, 2, ..., g-1\}$. The incentive compatibility (IC) requirement is,

$$u_i(v_i = A|\beta, n) \leq u_i(v_i = B|\beta, n)$$

which from the above expressions becomes,

$$\begin{aligned} & \varphi(n-1)[w(n+1)(1+k)] + [1-\varphi(n-1)][w(-n-1) + (1-w(-n-1))k] \\ \leq \ & \varphi(n-1)[w(n-1) + (1-w(n-1))k] + [1-\varphi(n-1)][w(-n+1)(1+k)] \end{aligned}$$

Rearranging in terms of k this constraint becomes, where $\overline{k}\,(n|g)$ represents equality,

$$\begin{aligned} \overline{k}\,(n|g) &= \frac{p^{n-1}[w(n-1) - w(n+1)] + (1-p)^{n-1}[w(-n+1) - w(-n-1)]}{p^{n-1}[w(n+1) + w(n-1) - 1] + (1-p)^{n-1}[1 - w(-n-1) - w(-n+1)]} \\ &\geq k \end{aligned}$$

Substituting in for w and differentiating, the derivative with respect to n has the following form,

$$\frac{d\overline{k}\,(n|g)}{dn} = \frac{\left[1 + \left(\frac{1-p}{p}\right)^{2g}\right]\left(\frac{1-p}{p}\right)^{g+n-1}(1-2p)^2(1-p)^{2n}p^{2(g+1)}\log\left[\frac{1-p}{p}\right]}{[f(g,p,n)]^2} < 0$$

where f is a certain function. By looking at the terms in the numerator and noticing that the denominator must be positive it can be seen that this expression is negative. Thus, the constraint corresponding to each value of n for β observers is tighter than all constraints for smaller values of n. If IC is satisfied when $n = g-1$ then it is satisfied for all smaller values of n. Using relationships for w derived in Lemma 3.2

this encompassing upper bound on k for a given g, denoted $\overline{k}_g$, is given by,

$$\overline{k}\left(g-1|g\right)=\overline{k}_g=\frac{\frac{\delta(1-p)^{g-2}}{p^{g+1}}\left[p^g-(1-p)^g\right]}{\left[p^{g-2}+(1-p)^{g-2}\right]-\frac{\delta(1-p)^{g-2}}{p^{g+1}}\left[p^g+(1-p)^g\right]} \tag{3.4}$$

A similar analysis for α observers must be performed. For this case consider $n\in\{0,1,2,...,g-1\}$. After observing α voter i faces the following two choices.

$$\begin{aligned}
u_i(v_i=A|\alpha,n) &= \varphi(n+1)[w(n+1)(1+k)] \\
&\quad +[1-\varphi(n+1)][w(-n-1)+(1-w(-n-1))k] \\
u_i(v_i=B|\alpha,n) &= \varphi(n+1)[w(n-1)+(1-w(n-1))k] \\
&\quad +[1-\varphi(n+1)][w(-n+1)(1+k)]
\end{aligned}$$

which are the same choices that faced a β observer but with different beliefs. The incentive compatibility requirement is,

$$u_i(v_i=A|\alpha,n)\geq u_i(v_i=B|\alpha,n)$$

which from the above expressions becomes,

$$\begin{aligned}
&\varphi(n+1)[w(n+1)(1+k)]+[1-\varphi(n+1)][w(-n-1)+(1-w(-n-1))k] \\
\geq\ &\varphi(n+1)[w(n-1)+(1-w(n-1))k]+[1-\varphi(n+1)][w(-n+1)(1+k)]
\end{aligned}$$

Rearranging in terms of k this constraint becomes, where $\underline{k}_*\left(n|g\right)$ represents equality,

$$\begin{aligned}
\underline{k}_*\left(n|g\right) &= \frac{p^{n+1}[w(n-1)-w(n+1)]+(1-p)^{n+1}[w(-n+1)-w(-n-1)]}{p^{n+1}[w(n+1)+w(n-1)-1]+(1-p)^{n+1}[1-w(-n-1)-w(-n+1)]} \\
&\leq k
\end{aligned}$$

Substituting in for w and differentiating, the derivative with respect to n has the following form,

$$\frac{d\underline{k}_*(n|g)}{dn} = \frac{-\left[1+\left(\frac{1-p}{p}\right)^{2g}\right]\left(\frac{1-p}{p}\right)^{g+n-1}(1-2p)^2(1-p)^{2(n-1)}\log\left[\frac{1-p}{p}\right]}{[h(g,p,n)]^2} > 0$$

where h is a certain function. By looking at the terms in the numerator and noticing that the denominator must be positive it can be seen that this expression is positive. Thus, the tightest constraint for all α observers is also for the greatest n, which is $n = g-1$. Once again using the relationships from Lemma 3.2 this encompassing lower bound on k for a given g, denoted $\underline{k}_{*g}$, is given by,

$$\underline{k}_*(g-1|g) = \underline{k}_{*g} = \frac{\frac{\delta(1-p)^g}{p^{g-1}}[p^{g-2}-(1-p)^{g-2}]}{[p^g+(1-p)^g] - \frac{\delta(1-p)^g}{p^{g-1}}[p^{g-2}+(1-p)^{g-2}]} \tag{3.5}$$

Now consider voters after the bandwagon has started. That is, after one of the candidates has gained a g vote lead. Without loss of generality assume that this has been candidate A. As these voters are expected to vote uninformatively the beliefs of subsequent voters are unaffected by their actions in equilibrium and so each will face the same decision problem (assuming, of course, that play is along the equilibrium path. This also explains why I need only consider $n = g$ as n can't increase past g). If one of these agents were to deviate and vote against the bandwagon then by the assumption of "sensible" beliefs, the beliefs of subsequent agents are below the bandwagon threshold and informative voting restarts until a g vote lead is again established by one of the candidates. Firstly, consider α observers. The utility payoffs of their two choices are as follows.

$$\begin{aligned} u_i(v_i = A|\alpha, g) &= \varphi(g+1) + k \\ u_i(v_i = B|\alpha, g) &= \varphi(g+1)[w(g-1) + (1-w(g-1))k] \\ &\quad +[1-\varphi(g+1)][w(-g+1)(1+k)] \end{aligned}$$

I will show that the constraint on k here is weaker than that of $\underline{k}_{*g}$, which was from an observer of α when A had a $(g-1)$ vote lead. Both of these voters are required to vote for A in equilibrium. I will show that, for any particular k, if the earlier voter has incentive to vote for A then so too does the voter after the bandwagon has started. The utilities of the two choices for the pre-bandwagon voter are given by,

$$\begin{aligned} u_i(v_i = A|\alpha, g-1) &= \varphi(g) + k \\ u_i(v_i = B|\alpha, g-1) &= \varphi(g)[w(g-2) + (1 - w(g-2))k] \\ &\quad +[1-\varphi(g)][w(-g+2)(1+k)] \end{aligned}$$

The following expressions are taken from the difference, $u_i(v_i = B|.) - u_i(v_i = A|.)$. Subscript 1 will denote the pre-bandwagon voter and subscript 2 the post-bandwagon voter. Firstly I will compare the non-k terms for these two voters. Δ reflects the pure informational incentive for each voter to deviate and vote for B.

$$\begin{aligned} \Delta_1 &= -\varphi(g)[1 - w(g-2)] + [1-\varphi(g)]w(-g+2) \\ \Delta_2 &= -\varphi(g+1)[1 - w(g-1)] + [1-\varphi(g+1)]w(-g+1) \end{aligned}$$

Using the expressions for φ and w, and relationships on w, this simplifies to,

$$\begin{aligned} [p^g + (1-p)^g]\Delta_1 &= -p^g \left(\frac{1-p}{p}\right)^{2g-2} \left(\frac{\delta}{p}\right) + (1-p)^g \frac{\delta}{p} \text{ , and} \\ [p^{g+1} + (1-p)^{g+1}]\Delta_2 &= -p^{g+1} \left(\frac{1-p}{p}\right)^{2g-1} \delta + (1-p)^{g+1}\delta \\ &= -p^g \left(\frac{1-p}{p}\right)^{2g-2} \left(\frac{\delta}{p}\right) p(1-p) + \left[(1-p)^g \frac{\delta}{p}\right] p(1-p) \\ &= p(1-p)[p^g + (1-p)^g]\Delta_1 \text{, which gives,} \\ \Delta_2 &= \frac{p^{g+1}(1-p) + (1-p)^{g+1}p}{p^{g+1} + (1-p)^{g+1}} .\Delta_1 \\ &< \Delta_1 \end{aligned}$$

Thus, for non-k terms the pre-bandwagon voter has greater incentive to deviate and vote B than does the post-bandwagon voter. Turning now to the k terms,

$$\begin{aligned}\psi_1 &= -1+\varphi(g)[1-w(g-2)]+[1-\varphi(g)]w(-g+2), \text{ and} \\ \psi_2 &= -1+\varphi(g+1)[1-w(g-1)]+[1-\varphi(g+1)]w(-g+1)\end{aligned}$$

Rearranging and simplifying produces,

$$\begin{aligned}[p^g+(1-p)^g][1+\psi_1] &= p^g\left(\frac{1-p}{p}\right)^{2g-2}\left(\frac{\delta}{p}\right)+(1-p)^g\left(\frac{\delta}{p}\right), \text{ and,} \\ \left[p^{g+1}+(1-p)^{g+1}\right][1+\psi_2] &= p(1-p)\left[p^g\left(\frac{1-p}{p}\right)^{2g-2}\left(\frac{\delta}{p}\right)+(1-p)^g\left(\frac{\delta}{p}\right)\right] \\ &= p(1-p)[p^g+(1-p)^g][1+\psi_1] \\ [1+\psi_2] &= \left[\frac{p^{g+1}(1-p)+(1-p)^{g+1}p}{p^{g+1}+(1-p)^{g+1}}\right][1+\psi_1] \\ &< [1+\psi_1] \\ \text{And so, } \psi_1 &> \psi_2\end{aligned}$$

Thus, the pre-bandwagon voter has greater incentive to deviate by voting B in terms of the k terms than does the post-bandwagon voter. Therefore, if the pre-bandwagon voter has incentive to follow the equilibrium strategy and vote A then so do α observers after the bandwagon has started. This implies the constraints imposed by α observers after the bandwagon has already started are satisfied if $\underline{k}_{*g}$ is satisfied.

Consider now the incentives of a β observer after the bandwagon has started. As with the α observers all such voters will face the same decision problem and there is only one constraint to consider, which shall be denoted by $\underline{k}_g$. Unlike post-bandwagon α observers, this condition will not be dominated by the condition from pre-bandwagon β observers. This is because the required equilibrium behavior of post-bandwagon β observers is the opposite of that from pre-bandwagon β observers. In fact this presents a lower bound on k rather than the upper bound imposed by pre-bandwagon β observers. The utilities from the two choices facing this voter are given as follows.

$$\begin{aligned} u_i(v_i = A|\beta, g) &= \varphi(g-1) + k \\ u_i(v_i = B|\beta, g) &= \varphi(g-1)[w(g-1) + (1 - w(g-1))k] \\ &\quad +[1 - \varphi(g-1)][w(-g+1)(1+k)] \end{aligned}$$

The incentive compatibility requirement is,

$$u_i(v_i = A|\beta, g) \geq u_i(v_i = B|\beta, g)$$

which from the above expressions becomes,

$$\begin{aligned} \varphi(g-1) + k &\geq \varphi(g-1)[w(g-1) + (1 - w(g-1))k] \\ &\quad +[1 - \varphi(g-1)][w(-g+1)(1+k)] \end{aligned}$$

Rearranging in terms of k this constraint becomes, where $\underline{k}_g$ represents equality,

$$k \geq \frac{p^{g-1}[w(g-1) - 1] + (1-p)^{g-1}w(-g+1)}{1 + p^{g-1}[w(g-1) - 1] - (1-p)^{g-1}w(-g+1)} = \underline{k}_g$$

Using relationships derived earlier for w this can be simplified to,

$$\underline{k}_g = \frac{\frac{\delta(1-p)^{g-1}}{p^g}[p^g - (1-p)^g]}{[p^{g-1} + (1-p)^{g-1}] - \frac{\delta(1-p)^{g-1}}{p^g}[p^g + (1-p)^g]} \tag{3.6}$$

Further investigations can show that $\underline{k}_g$ and $\underline{k}_{*g}$ cannot be ordered for all possible values of g and p. That is, for some pairs of g and p, $\underline{k}_g < \underline{k}_{*g}$ and for other values, $\underline{k}_g > \underline{k}_{*g}$. Thus both bounds will have to be compared to the upper bound, $\overline{k}_g$.

Now I will show that for a particular g, $\max\{\underline{k}_g, \underline{k}_{*g}\} < \overline{k}_g$ (which are stated in Equations 3.6,3.5, and 3.4, respectively), implying there exists a measurable region of k values for which this g-step bandwagon constitutes an equilibrium. I will prove this indirectly. Firstly I will show that $\overline{k}_{g+1} > \max\{\underline{k}_g, \underline{k}_{*g}\}$, where the subscript denotes the length of the corresponding bandwagon. Secondly, I show that $\overline{k}_g > \overline{k}_{g+1}$. Together these relationships imply that $\overline{k}_g > \max\left\{\underline{k}_g, \underline{k}_{*g}\right\}$.

A similar calculation to that which supplied Equation 3.4 produces the following,

$$\overline{k}_{g+1} = \frac{\left(\frac{\hat{\delta}}{p}\right)\frac{(1-p)^{g-1}}{p^{g+1}}[p^{g+1} - (1-p)^{g+1}]}{[p^{g-1} + (1-p)^{g-1}] - \left(\frac{\hat{\delta}}{p}\right)\frac{(1-p)^{g-1}}{p^{g+1}}[p^{g+1} + (1-p)^{g+1}]}$$

Where $\hat{\delta}$ is $w(-g)$ in a $(g+1)$-step bandwagon. Firstly I shall compare $\overline{k}_{g+1}$ with $\underline{k}_g$ by considering the numerators and denominators separately. I will show that the numerator is smaller and the denominator larger of $\underline{k}_g$ and thus as these terms are all positive it must be that $\underline{k}_g < \overline{k}_{g+1}$. Rearranging,

$$\text{numerator } \underline{k}_g = \delta\frac{(1-p)^{g-1}}{p^{g+1}}[p^{g+1} - p(1-p)^g]$$

This is of a similar form to the numerator of $\overline{k}_{g+1}$. Of the three components the final bracketed difference is obviously smaller in $\underline{k}_g$ as $p > \frac{1}{2}$. The middle term is the same so all that needs to be proven is that for all g the first term is also smaller. From the derivation of w earlier, $\delta = \hat{\delta}\left[\frac{1-\left(\frac{1-p}{p}\right)^{2g+2}}{1-\left(\frac{1-p}{p}\right)^{2g}}\right]$. As g is a positive integer it is immediate that $\hat{\delta} < \delta$, however the proof requires that $\frac{\delta}{\hat{\delta}} \leq \frac{1}{p}$. Rearranging the above expressions and labeling gives, $\eta = \frac{\delta}{\hat{\delta}} = \frac{1-\left(\frac{1-p}{p}\right)^{2g+2}}{1-\left(\frac{1-p}{p}\right)^{2g}}$. As $\frac{d\eta}{dg} < 0$, if $\frac{\delta}{\hat{\delta}} \leq \frac{1}{p}$ for $g = 1$ then it is true for all positive integers g. So for $g = 1$ this requirement becomes, $p \leq \frac{p^2}{2p^2-2p+1} \Rightarrow 0 \geq 2(p-\frac{1}{2})(p-1)$, which is true for $p \in (\frac{1}{2}, 1)$, the exact domain required. This proves that $\frac{\delta}{\hat{\delta}} \leq \frac{1}{p}$ and thus that the numerator of $\underline{k}_g$ is less than the numerator of $\overline{k}_{g+1}$.

Now turn to the denominators. The denominators can be written as follows.

$$\begin{aligned} \text{denominator } \underline{k}_g &= [p^{g-1} + (1-p)^{g-1}] - \delta\frac{(1-p)^{g-1}}{p^{g+1}}[p^{g+1} + p(1-p)^g], \text{ and} \\ \text{denominator } \overline{k}_{g+1} &= [p^{g-1} + (1-p)^{g-1}] - \left(\frac{\hat{\delta}}{p}\right)\frac{(1-p)^{g-1}}{p^{g+1}}[p^{g+1} + (1-p)^{g+1}] \end{aligned}$$

As the entire expression is positive and the first term in each expression is identical, to prove the claim requires,

$$\delta\frac{(1-p)^{g-1}}{p^{g+1}}[p^{g+1}+p(1-p)^g] \leq \left(\frac{\hat{\delta}}{p}\right)\frac{(1-p)^{g-1}}{p^{g+1}}[p^{g+1}+(1-p)^{g+1}]$$

Cancelling common terms simplifies the problem to proving that the following function is nonpositive for all positive integers g.

$$\tau(g)=\delta[p^{g+1}+p(1-p)^g]-\left(\frac{\hat{\delta}}{p}\right)[p^{g+1}+(1-p)^{g+1}]$$

Rearranging and substituting for δ, this becomes,

$$\tau(g)=p^{g+1}\hat{\delta}\left[\frac{1-\left(\frac{1-p}{p}\right)^{2g+2}}{1-\left(\frac{1-p}{p}\right)^{2g}}-\frac{1}{p}\right]+(1-p)^g p\hat{\delta}\left[\frac{1-\left(\frac{1-p}{p}\right)^{2g+2}}{1-\left(\frac{1-p}{p}\right)^{2g}}-\left(\frac{1-p}{p^2}\right)\right]$$

As g increases past 1 then $p^{g+1},(1-p)^g,\hat{\delta}$, and $\frac{1-(\frac{1-p}{p})^{2g+2}}{1-(\frac{1-p}{p})^{2g}}$ all decrease. Thus $\tau(g)$ also decreases in g for $g>1$. Combining this with the fact that $\tau(1)=0$ implies that $\tau(g)\leq 0$ for all positive integers g. Therefore, the denominator of $\underline{k}_g$ is larger than or equal to the denominator of $\overline{k}_{g+1}$. Combining this with the result about the numerators, this completes the proof that $\underline{k}_g<\overline{k}_{g+1}$, as required.

Now I shall prove that $\overline{k}_{g+1}>\underline{k}_{*g}$. Unfortunately this comparison is not as simple as comparing numerators and denominators. I will proceed by showing that the continuous function, $f(p,g)=\overline{k}_{g+1}-\underline{k}_{*g}>0$ over the interval $p\in(\frac{1}{2},1)$, and therefore the identity must be true for this domain.

After factoring out terms it proves beneficial to define the simplified problem of whether $f_1(p,g)>0$, where,

$$f^*(p,g) = \underline{k}_{*g}\cdot\frac{p^{g+2}}{(1-p)^{g-1}\left[1-\left(\frac{1-p}{p}\right)\right]}$$

$$= \frac{(1-p)[p^{g+1}-p^3(1-p)^{g-2}]}{\left[1-\left(\frac{1-p}{p}\right)^{2g}\right][p^g+(1-p)^g]-\left[1-\left(\frac{1-p}{p}\right)\right]\frac{(1-p)^g}{p^{g+2}}[p^{g+1}+p^3(1-p)^{g-2}]}$$

$$f^u(p,g) = \overline{k}_{g+1}\cdot\frac{p^{g+2}}{(1-p)^{g-1}\left[1-\left(\frac{1-p}{p}\right)\right]}$$

$$= \frac{[p^{g+1}-(1-p)^{g+1}]}{\left[1-\left(\frac{1-p}{p}\right)^{2g+2}\right][p^{g-1}+(1-p)^{g-1}]-\left[1-\left(\frac{1-p}{p}\right)\right]\frac{(1-p)^{g-1}}{p^{g+2}}[p^{g+1}+(1-p)^{g+1}]}$$

And, $f_1(p,g) = f^u(p,g) - f^*(p,g)$

Making the transformation, $\gamma = \frac{1-p}{p}$, simplifies these expressions further. This implies the additional relationships, $\frac{1}{\gamma} = \frac{p}{1-p}, p = \frac{1}{1+\gamma}$, and $1-p = \frac{\gamma}{1+\gamma}$. As $p \in (\frac{1}{2}, 1)$ then $\gamma \in (0,1)$. Making these substitutions produces,

$$f^*(\gamma,g) =$$

$$\frac{\left(\frac{\gamma}{1+\gamma}\right)\left[\left(\frac{1}{1+\gamma}\right)^{g+1}-\left(\frac{1}{1+\gamma}\right)^3\left(\frac{\gamma}{1+\gamma}\right)^{g-2}\right]}{(1-\gamma^{2g})\left[\left(\frac{1}{1+\gamma}\right)^g+\left(\frac{\gamma}{1+\gamma}\right)^g\right]-(1-\gamma)\gamma^g(1+\gamma)^2\left[\left(\frac{1}{1+\gamma}\right)^{g+1}+\left(\frac{1}{1+\gamma}\right)^3\left(\frac{\gamma}{1+\gamma}\right)^{g-2}\right]}$$

Multiplying by $\frac{(1+\gamma)^{g+2}}{(1+\gamma)^{g+2}}$ simplifies this expression to,

$$\begin{aligned} f^*(\gamma,g) &= \frac{\gamma[1-\gamma^{g-2}]}{(1-\gamma^{2g})\left[(1+\gamma)^2+\gamma^g(1+\gamma)^2\right]-(1-\gamma)\gamma^g(1+\gamma)^3(1+\gamma^{g-2})} \\ &= \frac{\gamma(1-\gamma^{g-2})}{(1+\gamma)^2\left[(1-\gamma^{2g})(1+\gamma^g)-(1-\gamma)\gamma^g(1+\gamma)(1+\gamma^{g-2})\right]} \end{aligned}$$

Making the same transformation to $f^u(p,g)$ produces,

$$f^u(\gamma,g) =$$

$$\frac{\left(\frac{1}{1+\gamma}\right)^{g+1} - \left(\frac{\gamma}{1+\gamma}\right)^{g+1}}{(1-\gamma^{2g})\left[\left(\frac{1}{1+\gamma}\right)^{g-1} + \left(\frac{\gamma}{1+\gamma}\right)^{g-1}\right] - (1-\gamma)\gamma^{g-1}(1+\gamma)^3\left[\left(\frac{1}{1+\gamma}\right)^{g+1} + \left(\frac{\gamma}{1+\gamma}\right)^{g+1}\right]}$$

Multiplying by $\frac{(1+\gamma)^{g+1}}{(1+\gamma)^{g+1}}$ simplifies this expression to,

$$\begin{aligned} f^u(\gamma,g) &= \frac{1-\gamma^{g+1}}{(1-\gamma^{2g})\left[(1+\gamma)^2+\gamma^{g-1}(1+\gamma)^2\right] - (1-\gamma)\gamma^{g-1}(1+\gamma)^3(1+\gamma^{g+1})} \\ &= \frac{1-\gamma^{g+1}}{(1+\gamma)^2\left[(1-\gamma^{2g+2})(1+\gamma^{g-1}) - (1-\gamma)\gamma^{g-1}(1+\gamma)(1+\gamma^{g+1})\right]} \end{aligned}$$

Combining these terms over a common denominator, and simplifying further, reduces the problem to proving $f_2(\gamma,g) > 0 \ \forall\gamma \in (0,1)$, where,

$$\begin{aligned} f_2(\gamma,g) &= (1+\gamma)^2 f_1(\gamma,g) \\ &= \frac{(\gamma-1)f_3(\gamma,g)}{(-1+\gamma^{2g})(1+\gamma^{1+g})(-\gamma^2+\gamma^{2g}-\gamma^{4+g}+\gamma^{2+3g})} \end{aligned}$$

And $f_3(\gamma,g)$ is given by the following,

$$f_3(\gamma,g) = -\gamma^2+\gamma^{2g}-\gamma^{1+g}-\gamma^{2+g}+\gamma^{1+2g}-\gamma^{3+2g}-\gamma^{4+2g}+\gamma^{2+3g}+\gamma^{3+3g}+\gamma^{2+4g}$$

Therefore, the initial problem of determining whether $f(p,g) > 0$ in the domain $p \in (\frac{1}{2},1)$ simplifies to the problem of whether $f_3(\gamma,g) < 0$ in the domain $\gamma \in (0,1)$ (as the final simplification involved division by a negative term). Considering the domain of γ, note that $\gamma^a < \gamma^b$ if $a > b \geq 1$. Therefore f_3 can be rewritten by grouping terms such that all bracketed pairs are negative for $\gamma \in (0,1)$ and thus the entire expression is negative in this domain (recall that $g \geq 1$).

$$\begin{aligned} f_3(\gamma, g) &= -(\gamma^2 - \gamma^{2g}) - (\gamma^{1+g} - \gamma^{1+2g}) - (\gamma^{2+g} - \gamma^{2+3g}) \\ &\quad -(\gamma^{3+2g} - \gamma^{3+3g}) - (\gamma^{4+2g} - \gamma^{2+4g}) \\ &< 0 \end{aligned}$$

Thus $f_3(\gamma, g) < 0$ for $\gamma \in (0, 1)$. Therefore, $f(p, g) > 0$ for $p \in (\frac{1}{2}, 1)$. It then follows that $\overline{k}_{g+1} > \underline{k}_{*g}$.

The proof that $\overline{k}_g > \overline{k}_{g+1}$ is similar to the proof that $\overline{k}_{g+1} > \underline{k}_{*g}$. Analogous steps reduces the current problem to proving that the function, $f_3^* > 0 \ \forall \gamma \in (0, 1)$, where,

$$f_3^*(\gamma, g) = \gamma + \gamma^{2g} - 2\gamma^{g+1} - \gamma^{g+2} + 2\gamma^{2g+1} - \gamma^{3g+1}$$

Factoring this expression,

$$\begin{aligned} f_3^*(\gamma, g) &= \gamma(1 - \gamma^g) + \gamma^{g+1}\left(\gamma^{g-1} - 1\right) + \gamma^{2g+1}(1 - \gamma^g) + \gamma^{g+2}\left(\gamma^{g-1} - 1\right) \\ &= (1 - \gamma^g)\gamma\left(1 + \gamma^{2g}\right) - \left(1 - \gamma^{g-1}\right)\gamma^{g+1}(1 + \gamma) \\ &> (1 - \gamma^g)\gamma\left(1 + \gamma^{2g}\right) - (1 - \gamma^g)\gamma^{g+1}(1 + \gamma) \\ &> (1 - \gamma^g)\gamma\left[1 + \gamma^{2g} - \gamma^g - \gamma^{g+1}\right] \\ &> 0 \end{aligned}$$

As $(1 - \gamma^g) > (1 - \gamma^{g-1})$, then $1 + \gamma^{2g} - \gamma^g - \gamma^{g+1} = (1 - \gamma^g) - \gamma^{g+1}(1 - \gamma^{g-1}) > 0$. Therefore, $f_3^*(\gamma, g) > 0$ for all $\gamma \in (0, 1)$ and $\overline{k}_g > \overline{k}_{g+1}$.

Obviously $\overline{k}_g > \overline{k}_{g+1}$ and $\overline{k}_{g+1} > \max\left\{\underline{k}_g, \underline{k}_{*g}\right\} \Rightarrow \overline{k}_g > \max\left\{\underline{k}_g, \underline{k}_{*g}\right\}$. Therefore, for any positive integer g, with "sensible" beliefs there exists a perfect Bayesian equilibrium for all $k \in [\max\{\underline{k}_g, \underline{k}_{*g}\}, \overline{k}_g]$, which is measurable and always exists.

I have shown that for any g there exists values of k that support this as an equilibrium. However, to prove the claim of the theorem the other direction must be proven: that for every value of k there exists a g such that a g-step bandwagon is an equilibrium. This proof requires two steps. The first step is already done. By proving

$\overline{k}_{g+1} > \underline{k}_g, \underline{k}_{*g}$ I have shown that the upper limit of the interval corresponding to a $(g+1)$-step bandwagon is never below the lower limit of the interval corresponding to a g-step bandwagon. Therefore, these intervals do not make jumps downwards such that some k values fall between consecutive intervals. I now show that as g approaches infinity the upper bound of these intervals, $\overline{k}_g$, approaches zero. Combining these two facts with the additional fact that for $g = 1$ there is no upper bound and an equilibrium exists for $k \geq \max\{\underline{k}_1, \underline{k}_{*1}\}$, it can be concluded that for any $k > 0$ there exists a corresponding g such that a g-step bandwagon is an equilibrium.

To prove $\lim\limits_{g\to\infty} \overline{k}_g = 0$ it is easier to deal with the original specification of $\overline{k}_g$,

$$\overline{k}_g = \frac{\varphi(g-2)[w(g-2)-1] + [1-\varphi(g-2)]w(-g+2)}{1-\varphi(g-2)[1-w(g-2)] - [1-\varphi(g-2)]w(-g-2)}$$

As $g \to \infty$, $\varphi(g-2), w(g-2) \to 1^-$. Therefore, $[1-\varphi(g-2)] \to 0$ and $w(-g+2) \to 0$. Combining these facts implies $[1-\varphi(g-2)]w(-g+2) \to 0$ and $\varphi(g-2)[1-w(g-2)] \to 0$, as $\varphi(g-2) < 1$. Therefore,

$$\lim_{g\to\infty} \overline{k}_g = \frac{0+0}{1-0-0} = 0$$

Assume now that there exists a $\tilde{k} > 0$ such that $\tilde{k} \notin \left[\max\left\{\underline{k}_g, \underline{k}_{*g}\right\}, \overline{k}_g\right]$ for all integers g. As $\overline{k}_g \to 0$ there must exist a g such that $\overline{k}_g < \tilde{k}$. Denote by $\tilde{g}$ the smallest g such that this is true ($\overline{k}_g$ exists for $g \geq 2$). If $\tilde{g} = 2$ then $\tilde{k} > \overline{k}_2 > \max\left\{\underline{k}_1, \underline{k}_{*1}\right\}$ and a one step bandwagon equilibrium exists. Consider now $\tilde{g} > 2$. By the definition of $\tilde{g}$ it must be that $\overline{k}_{(\tilde{g}-1)} \geq \tilde{k}$. By the proof above it is also true that $\max\left\{\underline{k}_{(\tilde{g}-1)}, \underline{k}_{*(\tilde{g}-1)}\right\} < \overline{k}_{\tilde{g}} < \tilde{k}$. Therefore, $\tilde{k} \in \left[\max\left\{\underline{k}_{(\tilde{g}-1)}, \underline{k}_{*(\tilde{g}-1)}\right\}, \overline{k}_{(\tilde{g}-1)}\right]$. A contradiction. Thus, the theorem must be true.

3.6.3 Proof of Corollary 3.1

{⇒sufficiency} This is the same as the symmetric case.

{⇒necessity} Let $(C_B, C_A) \in (0,1)$ such that the "informative vote lead" required for a bandwagon to begin for A is g_A, and for B it is g_B. Consider a β observer when

$n = g_A$. The two choices available to the voter are,

$$\begin{aligned} u_i(v_i &= A|\beta, g_A) = \varphi(g_A - 1) + k \\ u_i(v_i &= B|\beta, g_A) = \varphi(g_A - 1)[w(g_A - 1) + (1 - w(g_A - 1))k] \\ &\qquad + [1 - \varphi(g_A - 1)][w(-g_B + 1)(1 + k)] \end{aligned}$$

The incentive compatibility requirement is,

$$u_i(v_i = A|\beta, g_A) \geq u_i(v_i = B|\beta, g_A)$$

which from the above expressions becomes,

$$k \geq \frac{1 - \varphi(g_A - 1)[1 - w(g_A - 1)] - [1 - \varphi(g_A - 1)]w(-g_B + 1)}{\varphi(g_A - 1)[w(g_A - 1) - 1] + [1 - \varphi(g_A - 1)]w(-g_B + 1)} \tag{3.7}$$

The numerator must be strictly positive as the last two terms are mutually exclusive probabilities that do not cover all outcomes and must sum to less than one. Expanding the denominator and factoring gives,

$$\begin{aligned} Den &= \frac{p^{g_A - 1}}{p^{g_A-1} + (1-p)^{g_A-1}} \left[\frac{1 - \left(\frac{1-p}{p}\right)^{g_A + g_B - 1}}{1 - \left(\frac{1-p}{p}\right)^{g_A + g_B}} - 1 \right] \\ &\quad + \frac{(1-p)^{g_A-1}}{p^{g_A-1} + (1-p)^{g_A-1}} \left[\frac{1 - \left(\frac{1-p}{p}\right)}{1 - \left(\frac{1-p}{p}\right)^{g_A+g_B}} \right] \\ &= \frac{\left[1 - \left(\frac{1-p}{p}\right)\right] (1-p)^{g_A-1} \left[1 - \left(\frac{1-p}{p}\right)^{g_B}\right]}{\left[p^{g_A-1} + (1-p)^{g_A-1}\right] \left[1 - \left(\frac{1-p}{p}\right)^{g_A+g_B}\right]} \\ &> 0 \end{aligned}$$

Thus, if $(C_B, C_A) \in (0, 1)$ then IC requires $k > 0$ for a bandwagon equilibrium to exist.

Consider now $C_B = 0$ and $C_A < 1$. The case $C_A = 1$ and $C_B > 0$ is proven analogously. Equation 3.7 must still hold, except that now w can't be defined by

Lemma 3.1. Define the probability of victory in this case by w'. If the voter votes for B then either the boundary, g_A, is returned to or all voters vote informatively. If A is the better candidate then in either case A wins with probability one as the attainment of g_A starts a bandwagon in his favor, and informative voting leads, by the law of large numbers, to his vote share converging to p with probability one. Therefore $w'(g_A - 1) = 1$. For the denominator to be strictly positive it must be the case that $w'(-g_A + 1) > 0$, which is the probability that B wins when it has an unreachable threshold and A is one step (vote) away from his threshold.

Let g_B be an arbitrary and imaginary bound for a bandwagon on B to start, such that $g_A + g_B = x$. From Lemma 3.1, $\delta = w(-g_B + 1) > 1 - \left(\frac{1-p}{p}\right)$ for any finite g_B. So the probability that $n = g_A$ is reached before $n = -g_B$ is strictly less than $\left(\frac{1-p}{p}\right)$ (as one of the boundaries is reached with probability one).

Now starting from $n = -g_B$ consider an infinite sequence of expanding intervals in which the number of votes required for a bandwagon on B to start to be continually increasing. Let the next interval be of length $2x$, and the next of $4x$, and so on. From $n = -g_B$ the probability that $n = g_A$ is reached before -$(g_B + x)$ is $[1 - w_x(0)]$ (where the subscript denotes the length of the symmetric bandwagon). Otherwise $-(g_B + x)$ is reached. From there the probability that g_A is reached before $-(g_B + 3x)$ is $[1 - w_{2x}(0)]$, and so on. This gives the following expression for the probability that g_A is reached from $-g_B$.

$$
\begin{aligned}
P(g_A | -g_B) &= [1 - w_x(0)] + w_x(0)[1 - w_{2x}(0)] + w_x(0)\, w_{2x}(0)[1 - w_{4x}(0)] \\
&\quad + w_x(0)\, w_{2x}(0)\, w_{4x}(0)[1 - w_{8x}(0)] + \ldots
\end{aligned}
$$

In this sequence the term containing $[1 - w_y(0)]$ is multiplied by $\frac{[1-w_{2y}(0)]w_y(0)}{1-w_y(0)}$ to produce the following term, which can be simplified as follows,

$$
\frac{[1 - w_{2y}(0)]\, w_y(0)}{1 - w_y(0)} = \frac{\left[\frac{\left(\frac{1-p}{p}\right)^{2y} - \left(\frac{1-p}{p}\right)^{4y}}{1-\left(\frac{1-p}{p}\right)^{4y}}\right]}{\left[\frac{\left(\frac{1-p}{p}\right)^{y} - \left(\frac{1-p}{p}\right)^{2y}}{1-\left(\frac{1-p}{p}\right)^{2y}}\right]} \left[\frac{1 - \left(\frac{1-p}{p}\right)^{y}}{1 - \left(\frac{1-p}{p}\right)^{2y}}\right] = \frac{\left(\frac{1-p}{p}\right)^{y}}{1 + \left(\frac{1-p}{p}\right)^{2y}}
$$

As this ratio is decreasing in y, $P(g_A|-g_B)$ is strictly less than the sum of a convergent geometric progression with first term $[1-w_x(0)]$, and common ratio $\frac{\left(\frac{1-p}{p}\right)^x}{1+\left(\frac{1-p}{p}\right)^{2x}}$. This implies, $P(g_A|-g_B) < \frac{\left[\frac{\left(\frac{1-p}{p}\right)^x-\left(\frac{1-p}{p}\right)^{2x}}{1-\left(\frac{1-p}{p}\right)^{2x}}\right]}{1-\frac{\left(\frac{1-p}{p}\right)^x}{1+\left(\frac{1-p}{p}\right)^{2x}}} = \frac{\frac{\left(\frac{1-p}{p}\right)^x}{1+\left(\frac{1-p}{p}\right)^x}}{1-\frac{\left(\frac{1-p}{p}\right)^x}{1+\left(\frac{1-p}{p}\right)^{2x}}}$. For x increasing the denominator approaches one and the numerator approaches zero. As x depends on g_B, which is arbitrary, for any $\varepsilon > 0$ a g_B can be chosen such that $P(g_A|-g_B) < \varepsilon$.

From $n = g_A - 1$ the probability that g_A is reached is given by, $P(g_A|g_A-1) < \left(\frac{1-p}{p}\right) + \left[1-\left(\frac{1-p}{p}\right)\right] P(g_A|-g_B)$. As g_B is arbitrary it must be that $P(g_A|g_A-1) < 1$. So with strictly positive probability $n = g_A$ is never reached and, therefore, all voters vote informatively. By the law of large numbers the vote share for B in this case must converge to p with probability one. Therefore, with strictly positive probability B wins from $n = g_A - 1$ and so $w'(-g_A+1) > 0$. This implies that the denominator is strictly positive and so for IC to be satisfied and a bandwagon equilibrium to exist it must be that $k > 0$.

3.6.4 Proof of Corollary 3.3

It is easy to see that the $k > 0$ type voters face identical incentives to those in the symmetric case. As such, the analysis from that case carries over to this corollary. Thus, all that is required is to prove it is weakly optimal for the $k = 0$ type voters to follow the ABV strategy. I will separate this analysis into three parts. Firstly, if $n = 0$ then the previous analysis applies (as the vote will be informative) and the non-existence of bounds on k indicates that $v_i = s_i$ is optimal. Secondly, consider when $|n| \geq g$. This also follows immediately from the proof for $k > 0$ types when $|n| > g$. This proof also holds for $|n| = g$ in this case as votes against the leading candidate are uninformative for $k = 0$ type voters. Finally, suppose that $n < gg$. The utility choices facing voters in this situation are as follows.

$$\begin{aligned}
u_i(v_i = A|\alpha, n) &= \varphi(n+1)\, w(n+1) + [1-\varphi(n+1)]\, w(-n-1) \\
u_i(v_i = B|\alpha, n) &= \varphi(n+1)\, w(n) + [1-\varphi(n+1)]\, w(-n)
\end{aligned}$$

The equilibrium condition requires that $u_i(v_i = B|\alpha, n) \geq u_i(v_i = A|\alpha, n)$. Substituting for w and φ produces,

$$\begin{aligned}
\left(\frac{p^{n+1}}{p^{n+1}+(1-p)^{n+1}}\right)\delta\left(\frac{1-p}{p}\right)^{n+g} &\leq \left(\frac{(1-p)^{n+1}}{p^{n+1}+(1-p)^{n+1}}\right)\delta\left(\frac{1-p}{p}\right)^{-n-1+g} \\
&\Rightarrow \left(\frac{1-p}{p}\right)^{n} \leq 1
\end{aligned}$$

As $p > \frac{1}{2}$ this holds for all n and the equilibrium condition is satisfied. If the voter observes β then identical analysis produces the weaker condition $\left(\frac{1-p}{p}\right)^{n+2} \leq 1$, which must also hold.

Chapter 4 Vote Timing and Information Aggregation

Abstract

Drawing upon the results of Chapter 3 on sequential voting, and by deriving analogous results for simultaneous voting, I compare the information aggregation properties of the two processes. The conclusions confirm commonly held views about the front-loading of presidential primaries: that in tight races a simultaneous vote is preferred as it is more likely to lead to the selection of the "better" candidate, but in lopsided races a sequential vote is preferred. Strangely, the superior performance of sequential voting in lopsided races is precisely because bandwagons occur.

4.1 Introduction

In both the empirical and theoretical literatures, it is well documented and understood that the choice of voting rule can have a critical impact on the outcome of the electoral process. In a similar vein it is being increasingly documented that the timing of votes under any particular voting rule can also be critical to the outcome. This is nowhere more apparent than in the study of U.S. presidential primaries.[1]

To understand what this implies for politics in America, and more generally in any decision situation that employs sequential voting, it would be of interest to compare the information aggregation properties of this process with those of simultaneous voting. However, to this point there has been little theoretical work in this direction.[2]

The basic theoretical insight on this question, due to Dekel and Piccione (2000), is that in the standard voting model timing doesn't necessarily affect behavior or outcomes. They show that any informative symmetric equilibrium to the simultaneous voting game is also an equilibrium when voting is sequential. Therefore, the successful information aggregation results of Feddersen and Pesendorfer (1997) are applicable to the sequential voting environment, and thus in large electorates sequential and simultaneous voting are equivalent in their abilities to aggregate information.[3]

These findings do not, however, correlate well with the empirical observation that when voting is sequential, later voters are often influenced by the decisions of earlier voters and that timing does matter.[4] This influence usually takes the form of a "bandwagon," which is when voters later in the vote sequence follow the precedent of earlier voters and vote for the leading candidate. The dynamic phenomenon of

[1]This literature is too abundant to list. For a complete list of references see either of the book length treatments by Bartels (1988) or Morton and Williams (2000).

[2]While there has been a few papers looking at the properties of sequential voting and many looking at simultaneous voting, to the best of my knowledge the only direct comparison is made in Dekel and Piccione (2000). Morton and Williams (2000) compare simultaneous voting and two period voting in a specialized information environment. They examine primarily the coordination problems that arise in policy driven multiple candidate elections rather than the information aggregation issues that are the concern of the current paper.

[3]For common interest environments, as considered in this chapter, equivalence also holds for finite populations when voters are restricted to monotonic strategies (see Dekel and Piccione (2000) for details).

[4]See, for example, Bartels (1988).

bandwagons in the U.S. presidential nomination process has been increasingly of interest to researchers. The primary question of this interest is whether the sequential nature of the process, and its attendant disproportionate concentration on early primaries, leads to the selection of the best available candidate. Despite this attention our understanding of the merits of sequential versus simultaneous voting remains incomplete.[5] Basically the two competing viewpoints are as follows: when voting is sequential early voters may disproportionately influence the outcome more than is desirable in the interests of information aggregation, on the other hand simultaneous voting favors front runners to such an extent that many, potentially viable, candidates are ex-ante eliminated from consideration.

In an influential study of primaries, Bartels (1988) shows that these concerns are not idle ones. He argues that in several instances candidates were selected who otherwise would not have been had the voting been simultaneous. He further argues that these beneficial nominees (most notably Jimmy Carter in 1976) have had a significant impact on ultimate policy outcomes. This conclusion is not unique to Bartels as shown by the many attempts in recent years of candidates to manipulate, and legislators to reform, the nomination process.[6] The possibility for reform further heightens the necessity for a rigorous understanding of the different processes.

The leading suggestion for reform is that the nomination be decided sequentially by several regional primaries. Another alternative receiving attention is to hold one simultaneous national primary. Additionally, these reform efforts are not evolving in a static environment as due to the self interested selection of their own primary dates the states are 'front-loading' the process in such a way that an effective national primary

[5]The dilemma is summed up well by Morton and Williams (2000).

> "Front-loading seems to advantage well-known front-runner candidates. On the other hand, when primaries are drawn-out there is empirical evidence that voters learn information about candidates that allows them to make what may be more informed decisions. Yet drawn-out primaries give early voters an opportunity to perhaps influence the results of the sequential voting, early voters who are not representative of the general national party membership." (p. 31)

[6]See Morton and Williams (2000) for an account of these efforts.

is evolving. The conclusions of this chapter impact directly on these alternatives.

The results I present show that when the standard model is extended in a natural way to incorporate voters' desire to vote for the winning candidate the intuitions about the two processes are simultaneously correct and, consequently, equivalence no longer holds. If voting is sequential in a common interest environment then, as shown in Chapter 3, the dangers posed by bandwagons are always present. I derive here analogous results for simultaneous voting and show that, in contrast, the problems posed by this process only arise in specific, quantifiable circumstances. From these results I am then able to characterize the conditions when each negative force dominates and recommend when each of the processes should be used.

I find that in tight races, when there does not exist an overwhelming front runner, simultaneous voting is preferred (in the sense that the "better" candidate is selected more often). In contrast, in lopsided races, when such a front runner exists, sequential voting is better. These results coalesce well with the empirical findings described above and confirm the commonly held views about the front-loading of U.S. primaries. Significantly, they also imply that in certain circumstances the occurrence of bandwagons can have a positive impact on the quality of the winning candidate. These results hold as long as the voters' incentive to vote for the winner is nonzero and serve to show that the processes are equivalent only in the limiting case when this incentive is zero.

In Chapter 3, I assumed for tractability purposes an infinite voting population. To permit comparison, I make the same assumption here for simultaneous voting. This is not an unusual assumption, and is made in the hope that the limit results are a satisfactory proxy for large finite populations. However, the results for simultaneous voting appear to rely heavily on the fact that each voters pivot probability is exactly zero in the limit (this was not the case for the sequential voting results of Chapter 3). This raises concern about the relevance of these results as in finite populations the pivot probabilities are always strictly positive, even though they may become very small in large populations. To ease these concerns, I derive in the final section of this chapter the equilibria for a large but finite population when voters have a desire to vote for the winner and voting is simultaneous. I show that the limit results are, perhaps surprisingly, an appropriate approximation for large but finite populations.

It is important to note that while most of the evidence of sequential voting is from presidential primaries (and thus why this example is the focus of the chapter), the results presented here are applicable to any voting environment, such as voting in a boardroom or justices voting over supreme court rulings. Though for the most part voting is held simultaneously in these environments it is usually possible for it to instead be held sequentially. The results presented here indicate when and why such a change would be advantageous. Consider the example of a manager facing an investment decision when he is concerned with conforming behavior by his advisors (that may be due to risk aversion or mere sycophancy). The results reported here indicate the situations in which the manager would be best served eliciting votes, or opinions, sequentially rather than simultaneously to ensure the most profitable decision is made.

The remainder of the chapter is organized as follows. The following section presents the main results. It derives the equilibria for the two processes and compares their information aggregation properties. The third section contains the simultaneous voting results for when the population is finite. The final section concludes and suggests several directions for future research.

Related Literature

In a recent paper, Dekel and Piccione (2000) also compare the equilibrium properties of sequential and simultaneous voting. They consider a similar model of incomplete information but instead assume a finite population of voters and consider only the limit case of $k = 0$.[7] They show that when voters are restricted to using monotonic strategies[8] in a common interest environment the optimal equilibrium is

[7] Where k is as defined in Chapter 3 and represents each voter's desire to vote for the winning candidate.

[8] A strategy is monotonic if the probability of a voter voting for a certain alternative is higher when her private signal is in favor of this alternative.

independent of the timing of votes.[9] Therefore, the information aggregation capabilities of sequential and simultaneous voting are equivalent. In fact, this result can be generalized slightly by noting that when $\pi \in [1-p, p]$ equivalence holds without the restriction to monotonic strategies ("informative voting" fully aggregates information for all timing structures).[10]

In this chapter I consider these same questions. I find that with a large population and $k = 0$ informational equivalence holds more generally. I show that the information aggregation properties of sequential and simultaneous voting are equivalent for all prior beliefs and with no restrictions on the strategies used by voters. In fact, both mechanisms lead to full efficiency. In the expanded model, however, I show that equivalence breaks down. When voters have a desire to vote for the winning candidate ($k > 0$) the timing of votes matters and can have a critical impact on the quality of the electoral process. I am able to characterize when the different timing schemes are preferred and subsequently show that the processes are equivalent only in the limiting case when the desire to vote for the winning candidate is zero.

4.2 Results

The model used for these results is the same as used in Chapter 3. I will present three groups of results. The first will restate results from Chapter 3 about the existence of bandwagon equilibria in sequential voting. The second section will derive analogous results for simultaneous voting. And the third section will derive the main result of the chapter, the characterization and comparison of the information aggregation properties of these two processes.

[9]Dekel and Piccione (2000) also consider non-common interest environments. For simplicity and exposition I consider here only the common interest environment.

[10]Information is fully aggregated in the sense that the chosen alternative would not change if all private information were common knowledge. This is referred to as "full information equivalence" by Feddersen and Pesendorfer (1997).

4.2.1 Sequential Voting

The main result of Chapter 3 is that even with "sensible" beliefs BWV is an equilibrium for all nonzero values of k and that in these equilibria a bandwagon starts with probability one. Further, as k approaches zero the bandwagon takes longer and longer to begin. This requires a candidate to achieve a greater vote lead before all voters will support him regardless of their private signal. However, the crucial finding is that no matter how small is k a bandwagon will eventually begin.

Theorem 4.1 *[Chapter 3, Theorem 3.1] Suppose that beliefs are "sensible", $k > 0$, and that voting is sequential. Then $\exists\ (C_B, C_A) \in (0,1)^2$, where $C_B = 1 - C_A$, such that Bandwagon Voting by all voters constitutes a perfect Bayesian equilibrium.*

In this equilibrium all voters vote informatively until beliefs in favor of a certain candidate pass through a threshold. After this point is reached all subsequent voters "jump onto the bandwagon" of that candidate and vote for him regardless of their signals. Interestingly, the "informative vote lead" needed for the threshold to be passed can vary, depending on the parameters. This flexibility is consistent with observation and its absence had been the source of a primary criticism of previous attempts to explain bandwagons.

The equilibrium threshold increases towards one as the utility from voting for the winner approaches zero, and thus the "informative vote lead" needed to pass this threshold approaches infinity. The next theorem shows that CPV is still an equilibrium for $k = 0$ but only for the limiting case of CPV(0,1). This implies that a bandwagon will never begin and all voters will vote informatively. Further, it is only in this limit of $k = 0$ that CPV(0,1) is found to be an equilibrium strategy profile.

Theorem 4.2 *[Chapter 3, Theorem 3.2 and Corollary 3.1] Suppose that beliefs are "sensible" and voting is sequential. Then CPV(0,1) ("informative voting") by all voters is a perfect Bayesian equilibrium if and only if $k = 0$.*

These results imply that if voters are using the CPV strategy then a bandwagon should start if $k > 0$, and not otherwise. If a bandwagon starts, some information will not be aggregated into the group decision and the best candidate will not be selected with certainty. However, in the limiting case of $k = 0$ bandwagons will not start, instead "informative voting" will be used by all voters and, consequently, with probability one the best candidate will be selected.

4.2.2 Simultaneous Voting

To enable comparison some analogous results for when voting is simultaneous will now be derived. To maintain consistency with the results for sequential voting an infinite number of voters will continue to be assumed. In Section 4.3 I will show that the results derived here for an infinite population are a good approximation for the equilibria of a large but finite population.

Theorem 4.3 *Suppose voting is simultaneous and that $k > 0$. Then CPV(0,1) ("informative voting") by all voters is an equilibrium if and only if $\pi \in [1-p, p]$.*[11]

Proof: Recall that by the law of large numbers the correct candidate is selected by the majority with probability one if voting is informative. The updated beliefs by an α observer and a β observer are respectively given by $\varphi(0,\alpha)$ and $\varphi(0,\beta)$, where $\varphi(0,\alpha) = \frac{p\pi}{p\pi+(1-p)(1-\pi)}$ and $\varphi(0,\beta) = \frac{(1-p)\pi}{p\pi+(1-p)(1-\pi)}$. Now assuming everyone else is voting sincerely an α observer faces the following choices.

$$\begin{aligned} u(v &= A|\alpha, 0) = \varphi(0,\alpha)[1+k] + [1-\varphi(0,\alpha)]\,[1+0] \\ u(v &= B|\alpha, 0) = \varphi(0,\alpha)[1+0] + [1-\varphi(0,\alpha)]\,[1+k] \end{aligned}$$

Incentive compatibility requires,

$$u(v = A|\alpha, 0) \geq u(v = B|\alpha, 0)$$

[11]This equilibrium also exists for large finite populations. As the population becomes large then each voter's pivot probability approaches zero. Therefore, as k is strictly positive there always exists a population large enough such that the pivot probability is dominated by the voter's desire to vote for the winner.

This implies that $\varphi(0, \alpha) \geq \frac{1}{2}$. A similar analysis for a β observer produces the requirement that $\varphi(0, \beta) \leq \frac{1}{2}$. Substituting these requirements back into the original expressions for beliefs and rearranging for π produces the requirement that $1 - p \leq \pi \leq p$. ∎

Therefore, voters are prepared to vote informatively only if their prior beliefs are sufficiently close to neutral. Unlike informative voting in a bandwagon equilibrium, the informative voting evident here is motivated purely by the desire to vote for the winning candidate. When voting is simultaneous, a voter can't influence the choices of any other voters.[12] Thus, with an infinite population the pivot probability of all voters is zero and, consequently, any one voter can't influence the outcome of the election. As $k > 0$ each voter maximizes her utility by voting for the candidate she believes most likely to win. If priors are such that there is a highly favored front runner then a voter will deviate and vote for this front runner regardless of her private signal (assuming that all other voters are voting informatively). However, for priors close to neutral the private signal will determine which candidate the voter believes to be "better." Therefore, to maximize her probability of voting for the winner each voter will vote in accordance with her private signal, and "informative voting" is an equilibrium.

For large but finite populations this theorem is also true for $k = 0$.[13] However, with an infinite population the restriction on prior beliefs is sufficient but not necessary for "informative voting" to constitute an equilibrium. In fact, CPV(0,1) by all voters is supportable as an equilibrium for any prior beliefs as in the limit the pivot probabilities for all voters are zero.

Corollary 4.1 *Suppose voting is simultaneous and that $k = 0$. Then CPV(0,1) ("informative voting") by all voters is an equilibrium for all $\pi \in [0, 1]$.*

---

[12] Assuming voters do not have the ability to correlate their votes.

[13] This result is due to Austen-Smith and Banks (1996).

Proof: From the proof of Theorem 4.3,

$$u(v = A|\alpha, 0) = u(v = B|\alpha, 0) = u(v = A|\beta, 0) = u(v = B|\beta, 0) = 1$$

Which is independent of beliefs and always holds. Thus, by indifference, CPV(0,1) by all voters is an equilibrium for all prior beliefs.[14] ∎

This leaves the question of what constitutes an equilibrium for prior beliefs that are outside $[1-p, p]$ when $k > 0$. Immediately it can be seen, as described above, that informative voting by all voters cannot constitute an equilibrium as voters would deviate and support the front runner. The following theorem shows that this result is in fact much stronger. Not only is informative voting by all voters not an equilibrium, but in no equilibrium can there be any aggregation of information.

Theorem 4.4 *Suppose voting is simultaneous and that $k > 0$. If $\pi \notin [1-p, p]$ then in any equilibrium the outcome must be a constant or completely random. Thus the selection of the winning candidate is independent of the state of nature and private signals.*

Proof: For any voting profile, σ, there exists a sequence of expected vote shares as votes are counted. As there exists an infinite number of voters, by the law of large numbers the vote share in the actual vote sequence will converge to the expected share with probability one. Depending on the sequence of expected vote shares this implies that with probability one A wins, B wins, or a random draw decides the winner. Thus, for both states of the world the probability that either candidate wins can only take on one of the values, 0, $\frac{1}{2}$, or 1.

[14] The robustness of this result may appear doubtful as it holds only in the limit (when pivot probabilities are precisely zero). However, this equilibrium is in fact the limit of finite voter equilibria involving asymmetric strategies. Fey (1998), amongst others, point out that there is an equilibrium whereby enough voters vote uninformatively for the leading candidate such that the remaining voters have incentive to vote informatively. Thus for any particular prior the proportion of voters who vote informatively approaches one as the population gets large. And in the limit we have fully informative voting as described here.

Given state A is the true state, denote the probability that A wins by $P(A|A)$. Likewise denote the probability that A wins when B is the true state by $P(A|B)$. If $P(A|A) = P(A|B)$ then there may be an equilibrium. To prove the theorem I must prove there cannot be an equilibrium for $P(A|A) \neq P(A|B)$. Consider this in cases, relying on the fact that in the limit no individual voter can affect the result of the election. Assume that $\pi > p$. An identical argument proves the result for $\pi < 1 - p$.

Case 1: $P(A|A) = 1, P(A|B) \in \{0, \frac{1}{2}\}$. Then A wins with certainty whenever the true state is A. As $\varphi(\pi, \beta) > \frac{1}{2}$ the belief that the winner will be A is greater than $\frac{1}{2}$ regardless of the private signal. Thus all voters will vote for A (to receive k with greater probability). Thus, $P(A|B) = 1$ also and a contradiction is established.

Case 2: $P(A|A) = 0, P(A|B) \in \{\frac{1}{2}, 1\}$. In this case A loses with certainty whenever A is the true state. An identical argument to Case 1 then shows that $P(A|B) = 0$ and a contradiction is established.

Case 3: $P(A|A) = \frac{1}{2}, P(A|B) \in \{0, 1\}$. Voters are indifferent when the true state is A, and so base their vote choice on the expected outcome when B is the true state. If $P(A|B) = 0$, then voters strictly prefer B in state B and so strictly prefer B regardless of their signal. Thus all voters vote for B and $P(A|A) = 0$, which establishes a contradiction. Likewise for $P(A|B) = 1$, all voters vote for A and $P(A|A) = 1$ provides the necessary contradiction. ∎

Obviously then, any equilibrium for the simultaneous voting game in this domain does not aggregate information at all. Examples of such equilibria are for all voters to vote A regardless of their private signal, for all voters to vote B, or for all voters to completely randomize.[15]

[15] There exist many other equilibria but in all of them the outcome is uninformative. This is in spite of the possibility that in equilibrium infinite sets of voters can vote informatively! However, if this is the case then there must exist a countering set such that as a collective there is no informative voting. An example of such an equilibrium when the voters are represented by the set of positive integers, N, is the following.

Define $S = \{x | \forall x \in N \text{ s.t. } \exists y \in N \text{ where } x = 4y\}$,
And, $S' = \{w | \forall w \in N \text{ s.t. } \exists x \in S \text{ where } w = x + 1\}$.
Then equilibrium strategies are given by,
$\forall x \in S, \sigma_x(A|\alpha) = 1, \sigma_x(A|\beta) = 0$
$\forall y \in S', \sigma_y(A|\alpha) = 0, \sigma_y(A|\beta) = 1$
$\forall z \notin \{S \cup S'\}, \sigma_z(A|\alpha) = \sigma_z(A|\beta) = \frac{1}{2}$.

These results imply that for $k > 0$ there exists a discontinuity in the information aggregation properties of the simultaneous voting mechanism as the initial beliefs vary. For a tight race, where the difference in belief over the two candidates isn't too broad, simultaneous voting aggregates information completely. However, once one of the thresholds, p or $1 - p$, is passed, and the race becomes sufficiently lopsided, any equilibrium outcome does not aggregate information at all. The equilibrium that maximizes the probability the "better" candidate is selected, and also the Pareto dominant equilibrium, is for all voters to simply vote for the leading candidate. This equilibrium outcome confirms the fears of Schneider (1997, p. 734) when discussing the front-loading of U.S. presidential primaries,

> "...the 1996 contest also revealed the downside of front-loading. Candidates had to raise a lot of money early. That favored more-established figures such as Dole. Front-loading kept potentially strong candidates from running. And made it impossible for late starters to get in the race. 'I think we've probably made a mistake by front-loading these primaries,' GOP strategist Roger J. Stone said after last year's New Hampshire primary. 'It meant that good men like Jack Kemp and Colin Powell and Dick Cheney didn't make the race because they thought it was either unwinnable or too difficult.'"

4.2.3 Information Aggregation

The information aggregation properties of the simultaneous and the sequential voting games can now be compared. I find that when the utility from voting for the likely winner is nonzero the theoretical predictions are in line with empirical observation. Sequential voting is preferred if there exists a front runner who would otherwise dominate the race if voting were simultaneous. On the other hand, for close races simultaneous voting is preferred as under sequential voting the early voters have disproportionate influence and information aggregation is not complete. Critically, this inefficiency is not because early voters are unrepresentative of the whole voting population, as is commonly claimed to be the case in the presidential primaries literature, but rather because of the desire of later voters to support the winner. Further, in the limiting case when voters have no additional incentive to vote for the winner the two rules are equivalent. These results are stated in Proposition 4.1. Firstly, the following measure must be defined.

Definition 4.1 *For any pair of electoral processes, one is weakly preferred to another if it, ex-ante, selects the "better" candidate with a weakly higher probability. A process is strictly preferred to another if its probability of selecting the "better" candidate is strictly higher.*

In other words, an electoral process is preferred if it aggregates the private information better. For the purposes of comparison I assume that for each process the equilibrium played is that which maximizes the probability that the "better" candidate is selected.

Proposition 4.1 *Suppose that beliefs are "sensible" and let prior beliefs be given by equation 3.1, where m is an integer. Let g be the maximum integer such that a g-step BWV equilibrium exists.*

(a) $k > 0$. *If $m \in \{-1, 0, 1\}$ simultaneous voting is strictly preferred to sequential voting. However, if $1 < |m| < g$ sequential voting is strictly preferred to simultaneous voting. For $|m| \geq g$ sequential voting is weakly preferred to simultaneous voting.*

(b) $k = 0$. *The two rules are equivalent (full information aggregation).*[16]

Proof: Denote by P^{seq} and P^{sim} the maximum probabilities of selecting the "better" candidate under sequential and simultaneous voting, respectively. For each possible case the relationship between these two values needs to be established and then

[16]This is similar to the result of Dekel and Piccione (2000). They prove equivalence in this limit case ($k = 0$) for when the voting population is finite and voters are restricted to using monotonic strategies. See page 170 for a discussion of their result.

the proposition follows. Also, denote this same maximum probability for a g-step bandwagon equilibrium by $P^{bw}(g)$.

(a) Full information revelation under any electoral process requires an infinite number of voters to vote informatively, and for these voters to be decisive (decisive means that the majority of these informative voters determine the winner of the election). If voting is sequential then the arguments of Theorem 4.2 apply to the set of informative voters. Consequently members of this set have incentive to deviate and this can not be an equilibrium. Thus, for all m, $P^{seq} < 1$. For $m \in \{-1, 0, 1\}$ CPV(0,1) is an equilibrium of the simultaneous voting game and $P^{sim} = 1$. Therefore $P^{sim} > P^{seq}$.

If $1 < |m| < g$ and voting is simultaneous then all equilibria are uninformative. Thus, $P^{sim} = \pi(|m|)$ (when all voters vote for the leading candidate). However, if voting is sequential then a BWV equilibrium always exists. Consider the bandwagon equilibrium for $|m| < g$. Define $w(n)$ as the probability that a candidate will win from neutral priors if it is the "better" candidate and has an n "informative vote lead." In Chapter 3, I proved that this function has the following form, $w(n) = \frac{1-\left(\frac{1-p}{p}\right)^{n+g}}{1-\left(\frac{1-p}{p}\right)^{2g}}$. The probability of selecting the "better" candidate can then be written as follows.

$$
\begin{aligned}
P^{bw}(g) &= P(A|A) + P(B|B) \\
&= \varphi(m)w(m) + [1-\varphi(m)]\, w(-m) \\
&= \frac{p^m\left[1-\left(\frac{1-p}{p}\right)^{m+g}\right] + (1-p)^m\left[1-\left(\frac{1-p}{p}\right)^{g-m}\right]}{[p^m + (1-p)^m]\left[1-\left(\frac{1-p}{p}\right)^{2g}\right]} \\
&= \frac{p^m + (1-p)^m - \left(\frac{1-p}{p}\right)^{g}[(1-p)^m + p^m]}{[p^m + (1-p)^m]\left[1-\left(\frac{1-p}{p}\right)^{2g}\right]} \\
&= \frac{1-\left(\frac{1-p}{p}\right)^{g}}{1-\left(\frac{1-p}{p}\right)^{2g}} \\
&= w(0)
\end{aligned}
$$

Which, surprisingly, is independent of the particular value of m. Further calculations give the following.

$$\begin{aligned}
P^{bw}(g) &= w(0) = \frac{1-\left(\frac{1-p}{p}\right)^{g}}{1-\left(\frac{1-p}{p}\right)^{2g}} \\
&= \frac{p^{g}\left[p^{g}-(1-p^{g})\right]}{p^{2g}-(1-p)^{2g}} \\
&= \frac{p^{g}\left[p^{g}-(1-p^{g})\right]}{\left[p^{g}-(1-p)^{g}\right]\left[p^{g}+(1-p)^{g}\right]} \\
&= \frac{p^{g}}{p^{g}+(1-p)^{g}} \\
&= \pi(g)
\end{aligned}$$

As $|m| < g$ then $P^{seq} > P^{sim}$.

If $|m| \geq g$ it is still the case that $P^{sim} = \pi(|m|)$. Similarly if voting is sequential there still exists a bandwagon equilibrium and $P^{bw}(g) = \pi(|m|)$. However there may exist other equilibria for this process that are superior to the bandwagon equilibria (as, unlike for simultaneous voting, I have not solved for the complete set of equilibria under sequential voting). Consequently all that can be said is that $P^{seq} \geq \pi(|m|)$, and this implies that $P^{seq} \geq P^{sim}$.

(b) For $k = 0$ CPV(0,1) is an equilibrium under both voting processes. Thus all voters vote informatively, $P^{sim} = P^{seq} = 1$, and the proposition follows. ∎

As this result correlates well with empirical observation it may not be considered surprising. However, the cause is most surprising. Proposition 4.1 implies that the prospect of a bandwagon, in certain circumstances, can actually prove beneficial to the quality of candidate selected. This prospect is not entertained by the empirical literature that assumes that bandwagons are unambiguously bad for the quality of candidate selection.[17] This result is depicted in Figure 4.1 for when $k > 0$ and A is

[17] As noted earlier, the literature does point out that sequential voting can lead to the selection of superior candidates (see Bartels (1988)), though this is usually attributed to the mysterious notion of "momentum." The possibility that "bandwagon voting" is the cause of this early support for the underdog is not considered.

the front runner (the opposite case is symmetric to that drawn).

An understanding of why this is the case can be gained by considering the pivot probabilities of the early voters. When voting is simultaneous no voter can affect the vote choice of any other voter. Therefore, as the population grows large the pivot probability of each voter goes to zero. In contrast, when voting is sequential the early voters can influence the vote choices of later voters and their pivot probabilities remain strictly positive, even for large populations. In effect sequential voting permits votes to be correlated.

Therefore, when voting is sequential early voters may be enticed to vote informatively in the hope that this will lead to the selection of the "better" candidate, even if the existence of an overwhelming front runner implies that their choice is unlikely to win. In contrast, when voting is simultaneous no voter casts her ballot in the hope of affecting the result. Instead, all vote choices are made in the hope of supporting the winner, and it is only in tight races that this implies voters follow their private information and vote informatively. This follows from the fact that in tight races a voter's private information is her best indication of who is the "better" candidate.

It should be noted that the measure of preference stated in Definition 4.1 is not the only available measure. A natural alternative would be to prefer the electoral process that maximizes the sum of voters' utility. This differs from the current measure by incorporating voters' utility from conforming to the majority, thus placing a premium on consensus.[18] Though this measure, or possibly others, would permit a valid comparison of electoral processes, there are several reasons why Definition 4.1 was employed.

[18]Aside from the arguments for Definition 4.1 presented in the following paragraph, employing the sum of utilities in an infinite population is problematic. The difficulty is finding an informative yet noncontroversial measure of this aggregate utility. The two natural suggestions, unfortunately, are both inappropriate. Firstly, summing utilities over an infinite population is uninstructive (as it always gives ∞). Secondly, calculating average utility is similarly uninstructive. For example, in any bandwagon equilibrium almost all voters conform to the majority decision, regardless of the length of the bandwagon. Thus the average utility is always $\pi(g) + k$ and the fact that additional voters will not conform to the majority in longer bandwagons is lost even though their contribution to the quality of candidates is incorporated.

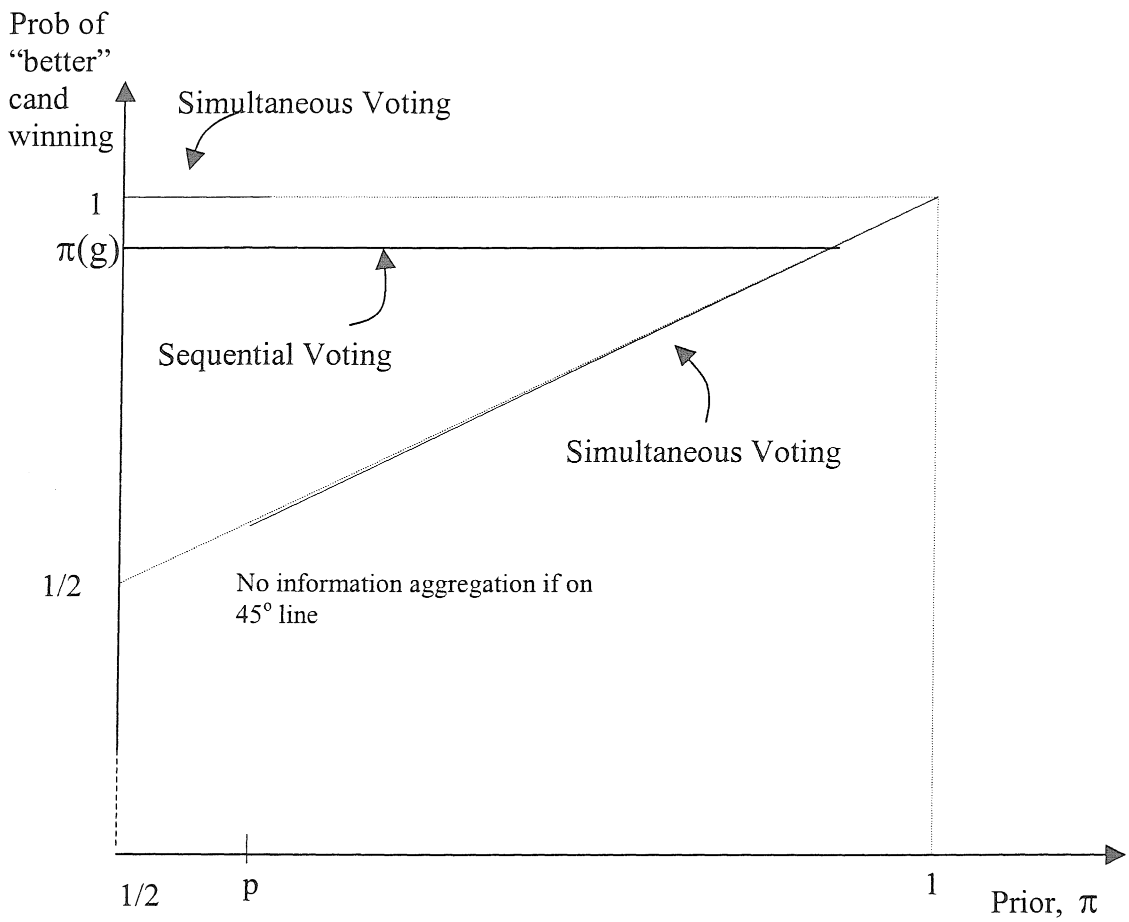


Figure 4.1: Sequential vs. Simultaneous Voting (Optimal Equilibria)

First, it focuses attention on the information aggregation capabilities of the voting mechanism under different timing schemes. Second, this would be the natural objective if the choice over timing schemes is made by an individual or group outside of the voting population. For example, a committee chair determining how an agenda will be voted on may care only about the outcomes and not the satisfaction committee members gain by being in a majority. Similarly, a manager soliciting votes or opinions on an investment project is most likely to be interested only in the viability of the project. More powerfully, drafters of constitutions are unlikely to be concerned with the additional conformist motivations, and indeed may actively attempt to circumvent them, when determining how decisions will be made that affect the entire population (e.g., when deciding voting rules and processes for legislators or justices). Finally, if k is interpreted in a distributional sense then Definition 4.1 provides a measure of social welfare. Earlier I indicated that one rationale for k was as a payoff from the winning candidate to his supporters. If this is the case then the receipt of k does not add to social welfare as it is a pure redistribution of wealth from nonvoters or voters in the minority (such a specification is equivalent to the current model as only the difference in utility from being in the majority or minority matters to the decisions of voters). Therefore, in this case, social welfare is maximized by maximizing the probability that the "better" candidate is selected.

Finally, it should be noted that the strategies used to prove Proposition 4.1, "informative voting" and Bandwagon voting, are both monotonic (and symmetric).[19] This implies that the informational equivalence of sequential and simultaneous voting breaks down even if voter strategies are restricted to be monotonic, as imposed by Dekel and Piccione (2000). Consequently, the breakdown of equivalence found here for common interest environments and an infinite population cannot be attributed to the expansion of the strategy sets available to voters. Instead we must conclude that equivalence breaks down in this environment because voters are driven by a desire to vote for the winning candidate.

[19]See footnote 8 on page 170 for a definition of monotonic strategies.

4.3 Simultaneous Voting with a Finite Population

The proofs of the results in Section 4.2.2 for simultaneous voting rely heavily on the fact that the pivot probabilities for all voters are identically equal to zero. This reliance is potentially concerning as in large but finite populations, which infinite populations are meant to proxy, the pivot probabilities are always strictly positive, even though they may be very small. Therefore, one may ask whether the results of the infinite voter model carry over, at least in approximate form, to a large but finite population. In this section I derive analogous results for finite populations and show that the results of Section 4.2.2 are indeed good approximations for voter behavior. I will focus here on equilibria when $k > 0$. The asymptotic properties of symmetric equilibria in the standard case when $k = 0$ have been characterized by Feddersen and Pesendorfer (1998). They showed that asymptotic efficiency can be achieved and therefore the large finite population results correspond exactly to the infinite population results derived in the previous section.

4.3.1 Finite Population Equilibria

In this section I will again employ the model of Chapter 3, with several alterations. Firstly, I assume an arbitrary population of $2n + 1$ voters for some integer n.[20] Definition 3.1 collapses to the standard notion of majority rule (i.e., the most votes wins). I will not require prior beliefs to be "sensible," instead they may take on any value in the interval $[0, 1]$. I will consider the following symmetric strategies: for all i, $\sigma_i(\alpha, \emptyset) = q_\alpha$ and $\sigma_i(\beta, \emptyset) = 1 - q_\beta$ (recall that $\sigma_i \in [0, 1]$ is the probability of voting for candidate A). The values q_α and q_β are the probabilities a voter votes sincerely given their private signal $s_i \in \{\alpha, \beta\}$. I will use the notation that $\sigma = \{q_\alpha, q_\beta\}$ implies $\sigma_i(\alpha, \emptyset) = q_\alpha$ and $\sigma_i(\beta, \emptyset) = 1 - q_\beta$ for all i.

If A is the true state then the probability a random voter votes for A, denoted by

[20]This implies that the population is always odd. This is for notational convenience only and the results hold for even populations also.

γ_A, is given by

$$\gamma_A = pq_\alpha + (1-p)(1-q_\beta)$$

Likewise, if B is the true state then the probability that a random voter votes for B, denoted by γ_B is given by

$$\gamma_B = pq_\beta + (1-p)(1-q_\alpha)$$

As the vote history is empty for all voters it will be omitted from all expressions. This allows a voter's posterior beliefs to be denoted by $\varphi(s_i)$.

As is standard in the finite population literature I will focus on symmetric equilibria.[21] If all other voters are using the identical strategy described by $\sigma = \{q_\alpha, q_\beta\}$ then the utility for each choice faced by the voter is given by the following:

$$\begin{aligned} u(v = A|s_i, \sigma) &= \varphi \sum_{j=n}^{2n} \binom{2n}{j} \gamma_A^j (1-\gamma_A)^{2n-j} \\ &+ (1-\varphi) \sum_{j=n+1}^{2n} \binom{2n}{j} \gamma_B^j (1-\gamma_B)^{2n-j} \\ &+ \left[\varphi \sum_{j=n}^{2n} \binom{2n}{j} \gamma_A^j (1-\gamma_A)^{2n-j} \right] k \\ &+ \left[(1-\varphi) \sum_{j=0}^{n} \binom{2n}{j} \gamma_B^j (1-\gamma_B)^{2n-j} \right] k \end{aligned}$$

[21] This focus is typically for tractability purposes as there would most likely exist many asymmetric equilibria. However, with voters also possessing a desire to vote for the winner the willingness of voters to behave asymmetrically may be restricted. I showed earlier in this chapter that in the limit asymmetric equilibria can perform no better than symmetric equilibria. How strong this restriction is in a finite population is an intriguing open question, and one I hope to investigate in the future.

$$
\begin{aligned}
u(v=B|s_i,\sigma) &= \varphi \sum_{j=n+1}^{2n} \binom{2n}{j} \gamma_A^j (1-\gamma_A)^{2n-j} \\
&\quad + (1-\varphi) \sum_{j=n}^{2n} \binom{2n}{j} \gamma_B^j (1-\gamma_B)^{2n-j} \\
&\quad + \left[\varphi \sum_{j=0}^{n} \binom{2n}{j} \gamma_A^j (1-\gamma_A)^{2n-j} \right] k \\
&\quad + \left[(1-\varphi) \sum_{j=n}^{2n} \binom{2n}{j} \gamma_B^j (1-\gamma_B)^{2n-j} \right] k
\end{aligned}
$$

Unless necessary for exposition, the arguments s_i and σ will be omitted hereafter. The utility difference between these two choices can now be written as (noting that $\binom{2n}{j} = \binom{2n}{2n-j}$),

$$
\begin{aligned}
\Phi(s_i) &= u(v=A|s_i) - u(v=B|s_i) \\
&= \varphi(s_i) \binom{2n}{n} \gamma_A^n (1-\gamma_A)^n - (1-\varphi(s_i)) \binom{2n}{n} \gamma_B^n (1-\gamma_B)^n \\
&\quad + k\varphi(s_i) \sum_{j=0}^{n-1} \binom{2n}{j} \left[\gamma_A^{2n-j} (1-\gamma_A)^j - \gamma_A^j (1-\gamma_A)^{2n-j} \right] \\
&\quad - k(1-\varphi(s_i)) \sum_{j=0}^{n-1} \binom{2n}{j} \left[\gamma_B^{2n-j} (1-\gamma_B)^j - \gamma_B^j (1-\gamma_B)^{2n-j} \right] \\
&= \Delta(s_i) + k\psi(s_i)
\end{aligned}
$$

As in the previous sections, I will use as my efficiency measure the probability that the "better" candidate is selected, and focus on the most efficient outcomes. To this end the following two results will prove useful in simplifying later analysis. The first lemma states the well known result that in voting games there always exist equilibria in which no information is aggregated. Denote by $P_n(\sigma)$ the probability that the "better" candidate is selected given strategy profile σ and population size $2n+1$.

Lemma 4.1 *The strategies* $\sigma' = \{0,1\}$ *and* $\sigma'' = \{1,0\}$ *are equilibria for all* n *and* $\pi \in [0,1]$. *Further,* $\forall n$, $\max\{P_n(\sigma'), P_n(\sigma'')\} = \max\{1-\pi, \pi\}$.

Proof: For strategy $\sigma', \Delta(\alpha) = \Delta(\beta) = 0$, and $\psi(\alpha) = \psi(\beta) = -1$. Therefore $\Phi(\alpha) = \Phi(\beta) = -k$, and σ is an equilibrium. The proof for σ'' is identical. ∎

The equilibria described in this lemma require that all voters ignore their private information and pool their votes onto the one candidate. Consequently, no information is aggregated. The existence of these equilibria places a lower bound on the efficiency of the voting mechanism. They imply that there will always exist equilibria in which the voting mechanism leads to no worse outcomes than could be achieved without its use. Surprisingly, these equilibria are not dominated by all equilibria in which information is aggregated within the voting mechanism. This bound allows these somewhat peculiar equilibria to be dismissed immediately as they cannot be optimal. It should be noted that the equilibria of Lemma 4.1 are strict if $k > 0$. In the standard model ($k = 0$) voters are indifferent in these uninformative equilibria as their action cannot affect their utility. However, when $k > 0$ such arguments do not apply as all voters strictly prefer to follow the equilibrium strategy in order to secure a payoff from voting for the certain winner.

The next lemma characterizes the strategy profiles which are dominated by the bound in Lemma 4.1. Surprisingly, in these profiles some information may be aggregated within the voting mechanism. However, the lemma implies that the information must be aggregated in such a way that outcomes are worse than if no information is aggregated at all. For any $\sigma = \{q_\alpha, q_\beta\}$ denote by $P(A|A, \sigma)$ and $P(B|B, \sigma)$ the probabilities that A and B win, respectively, given they are the "better" candidate.

Lemma 4.2 *If* $q_\alpha \leq 1 - q_\beta$ *then* $\sigma = \{q_\alpha, q_\beta\}$ *implies* $P_n(\sigma) \leq \pi$ *for all* n. *If* $q_\alpha < 1 - q_\beta$ *then* $P_n(\sigma) < \pi$.

Proof: $q_\alpha \leq 1 - q_\beta$ implies that $P(A|B, q_\alpha, q_\beta) \geq P(A|A, q_\alpha, q_\beta)$ and $P(B|B, q_\alpha, q_\beta) \leq P(B|A, q_\alpha, q_\beta)$. As $P(B|A) = 1 - P(A|A)$ and $P(B|B) = 1 - P(A|B)$ then $P(B|A) > P(B|B) \Leftrightarrow 1 - P(A|A) > 1 - P(A|B) \Leftrightarrow P(A|A) <$

$P(A|B)$. So WOLOG suppose that $P(B|A) > P(B|B)$. $P_n(\sigma)$ can now be written:

$$\begin{aligned} P_n(\sigma) &= \pi P(A|A) + (1-\pi) P(B|B) \\ &= \pi[1 - P(B|A)] + (1-\pi) P(B|B) \\ &\leq \pi[1 - P(B|B)] + (1-\pi) P(B|B) \\ &\leq \pi + P(B|B)(1-2\pi) \\ &\leq \max[1-\pi, \pi] \end{aligned}$$

And this inequality is strict if $q_\alpha < 1 - q_\beta$. ∎

Combining these two lemmas implies that, for all π and n, in any optimal equilibria it must be that $q_\alpha \geq q_\beta$. Therefore, from now on I shall focus on this case. The next lemma characterizes the possible equilibria for when both α and β observers play mixed strategies.

Lemma 4.3 *If $\sigma = \{q_\alpha, q_\beta\}$ is an equilibrium such that $q_\alpha, q_\beta \in (0,1)$ then $P_n(\sigma) < \pi$ for all n.*

Proof: If $q_\alpha < 1 - q_\beta$ then the result follows from Lemma 4.2. So suppose that $q_\alpha \geq 1 - q_\beta$. Then $P(B|A) \leq P(B|B)$ and $P(A|A) \geq P(A|B)$. Recall that $\Phi = \Delta + k\psi$ and $\varphi(\alpha) \neq \varphi(\beta)$. Therefore, if both Δ and ψ are increasing in φ, with at least one strictly increasing, then $\Phi(\alpha) \neq \Phi(\beta)$ and voters can't mix for both signals in an equilibrium. Firstly, $\frac{\partial \Delta}{\partial \varphi} = \binom{2n}{n}\gamma_A^n(1-\gamma_A)^n + \binom{2n}{n}\gamma_B^n(1-\gamma_B)^n > 0$. Secondly,

$$\begin{aligned} \frac{\partial \psi}{\partial \varphi} &= \sum_{j=0}^{n-1}\binom{2n}{j}\left[\gamma_A^{2n-j}(1-\gamma_A)^j - \gamma_A^j(1-\gamma_A)^{2n-j}\right] \\ &\quad + \sum_{j=0}^{n-1}\binom{2n}{j}\left[\gamma_B^{2n-j}(1-\gamma_B)^j - \gamma_B^j(1-\gamma_B)^{2n-j}\right] \\ &\geq 0 \text{ as } P(A|A) \geq P(A|B) \text{ and } P(B|A) \leq P(B|B) \end{aligned}$$

∎

I now turn to strategy profiles in which voters use a mixed strategy for at most one signal. I will consider two cases: tight races in which $\pi \in [1-p, p]$, and lopsided races in which $\pi > p$ or $\pi < 1-p$. Throughout I will assume A to be the ex-ante preferred candidate. This is without loss of generality as the case when B is ex-ante preferred is identical with only the candidate labels interchanged.

Theorem 4.5 *Suppose that $\pi \in [1-p, p]$, voters mix for at most one type of signal, and $q_\alpha \geq 1 - q_\beta$. For all integers n there exists at least the following five symmetric equilibria:*

1. $q_\beta = 0, q_\alpha = 1$ *(all voters vote A)*
2. $q_\beta = 1, q_\alpha = 0$ *(all voters vote B)*
3. $q_\beta = 1, q_\alpha = 1$ *(all voters vote informatively)*
4. $q_\beta = \tilde{q}_\beta \in (0,1), q_\alpha = 1$ *(β observers mix, α observers vote A)*
5. $q_\beta = 1, q_\alpha = \tilde{q}_\alpha \in (0,1)$ *(β observers vote B, α observers mix)*

Proof: 1. and 2. As proven in Lemma 4.1.

3. For $\sigma = \{1, 1\}$,

$$\begin{aligned}
\psi(s_i) &= \varphi(s_i) \sum_{j=0}^{n-1} \binom{2n}{j} \left[p^{2n-j}(1-p)^j - p^j (1-p)^{2n-j}\right] \\
&\quad - (1-\varphi(s_i)) \sum_{j=0}^{n-1} \binom{2n}{j} \left[p^{2n-j}(1-p)^j - p^j (1-p)^{2n-j}\right] \\
&= (2\varphi(s_i) - 1) \sum_{j=0}^{n-1} \binom{2n}{j} \left[p^{2n-j}(1-p)^j - p^j (1-p)^{2n-j}\right] \\
&> 0 \text{ if } \varphi(s_i) > \frac{1}{2} \text{ as } p > \frac{1}{2} \\
&= 0 \text{ if } \varphi(s_i) = \frac{1}{2} \\
&< 0 \text{ if } \varphi(s_i) < \frac{1}{2}
\end{aligned}$$

And,

$$\begin{aligned}
\Delta(s_i) &= \varphi(s_i)\binom{2n}{n}p^n(1-p)^n - (1-\varphi(s_i))\binom{2n}{n}p^n(1-p)^n \\
&= \binom{2n}{n}p^n(1-p)^n(\varphi(s_i) - (1-\varphi(s_i))) \\
&= \binom{2n}{n}p^n(1-p)^n(2\varphi(s_i) - 1)
\end{aligned}$$

Therefore:

$$\begin{array}{l}
\Delta(s_i) > 0 \text{ if } \varphi(s_i) > \frac{1}{2} \\
\Delta(s_i) = 0 \text{ if } \varphi(s_i) = \frac{1}{2} \\
\Delta(s_i) < 0 \text{ if } \varphi(s_i) < \frac{1}{2}
\end{array}$$

Consequently, as $\pi \in [1-p, p]$, $\Phi(\alpha) \geq 0$ and $\Phi(\beta) \leq 0$. Thus $\sigma = \{1,1\}$ is an equilibrium.

4. If $\sigma = \{1,0\}$ then Lemma 4.1 showed that $\Delta(s_i) = 0$ and $\psi(s_i) = 1$. By the continuity of Δ and ψ, and the arguments of 3 above, there exists at least one $\tilde{q}_\beta \in (0,1)$ such that $\sigma' = \{1, \tilde{q}_\beta\}$ is an equilibrium.

5. Analogous to 4. ∎

These results are depicted graphically in Figures 4.2a and 4.2b for when five symmetric equilibria exist. The different lines represent the different components of voter utility. For simplicity I have drawn ψ as being linear. This is the case when there are three voters, but will not be so generally (though it can be shown that it is always strictly monotonic in these environments).

The addition of the voters' desire to vote for the winner alters the standard analysis by expanding the set of equilibria. For prior beliefs in this range there would only exist three symmetric equilibria if $k = 0$. Equilibria 4 and 5 would no longer exist. These equilibria vary continuously in k and for small k they are arbitrarily close to equilibria 1 and 2, respectively. Thus the inclusion of a desire to vote for the winning candidate adds two inefficient equilibria to the voting game. It is worth noting that the optimal equilibria is number 3, in which all voters vote informatively and all private information is aggregated into the group decision. Thus, the optimal

Key: $\Delta \equiv$ ...., $\psi \equiv$ ----------,
$\Phi = u(v = A) - u(v = B) = \Delta + k\psi \equiv$ ________

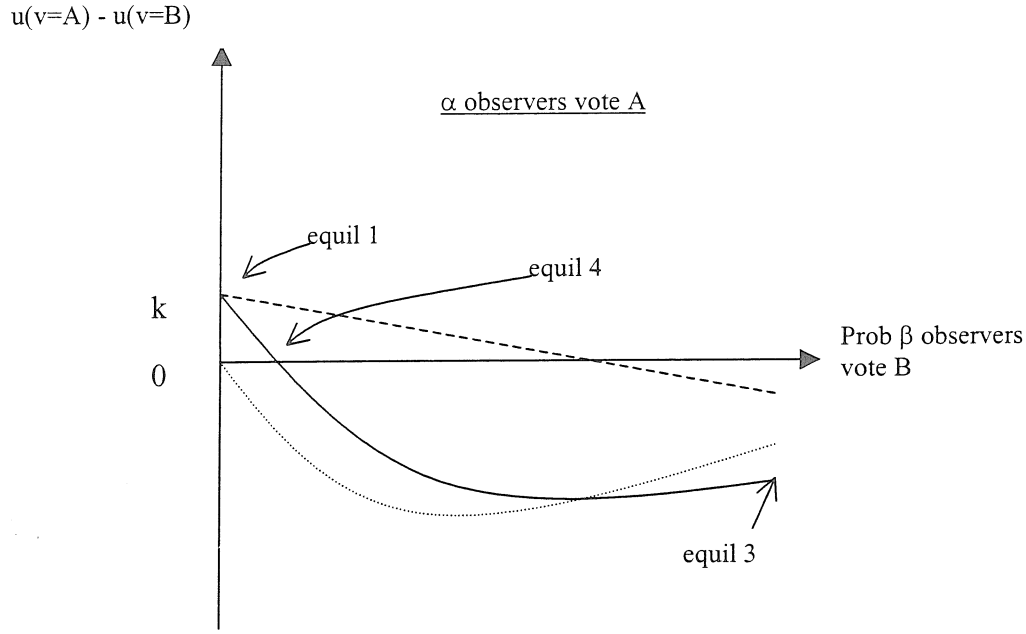


Figure 4.2a: $\pi \in [1-p,p]$

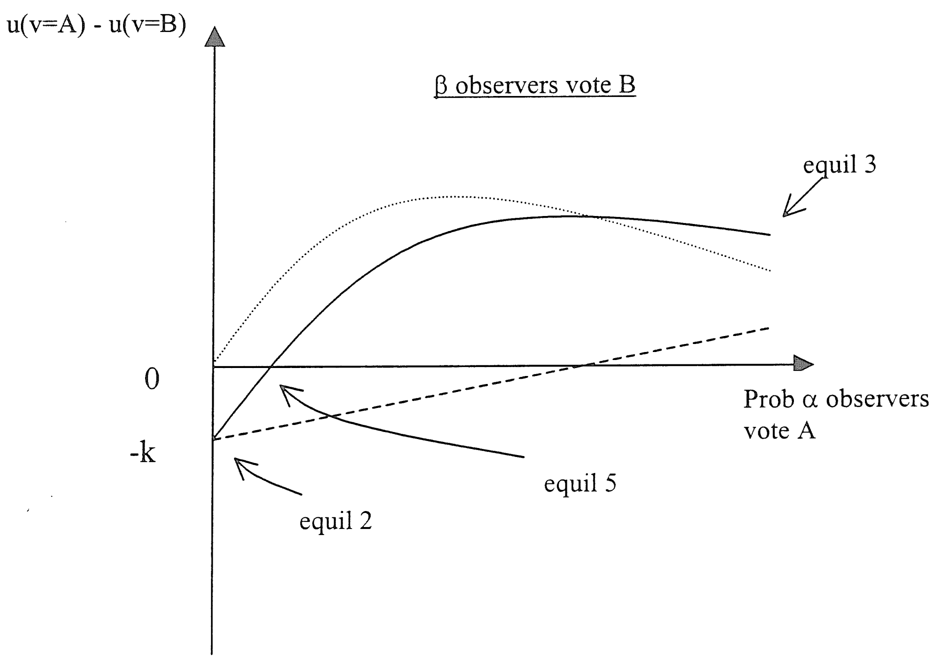


Figure 4.2b: $\pi \in [1-p,p]$

Figure 4.2: Finite Population Equilibria in Tight Elections

equilibria found in Section 4.2.2 for an infinite population when the election is tight also exists when the population is finite. This will be discussed more formally in Section 4.3.2.

The mixing equilibria are brought about when voters trade off their informational incentives with their desire to be in the winning majority. The existence of these equilibria are an indication of the effects a desire to vote for the winner has on voting environments. In, for example, equilibrium 4, β observers are induced to sometimes vote A even though they strongly believe that B is the "better" candidate when they are pivotal. They are induced to do this because the mixing of other voters makes A a more likely winner, and the β observers wish to support this likely winner. The inefficiency of this equilibrium is institutionalized as the voters' beliefs are self-reinforcing: A is more likely to win because β observers vote for him, and β observers vote for him because he is more likely to win. Significantly, this interdependence can result in inefficient equilibria even in an environment that is capable of inducing complete efficiency.

The interdependence of these beliefs is also striking because it requires a precise balance to be maintained in equilibrium. As such it may be questioned whether these mixing equilibria are 'stable' with respect to small perturbations and best response dynamics. From the figures it can be seen that the two mixing equilibria are in fact not stable but that the pure strategy equilibria are stable.[22] Consider, for example, equilibrium 4. If the probability that β observers voted B increased slightly, it would then be strictly in the interests of each β observer to vote B and they would no longer mix. This implies that a small increase in this probability would result in even greater increases, and for behavior to move away from the equilibrium level. Therefore, the two mixing equilibria drawn in the figures are not stable.

[22]This is proven formally in Proposition 4.2. Despite its frequent application, there does not seem to exist a standardized definition of 'stability' with respect to best response dynamics. I will suppose that best response dynamics imply, for some strategy profile $\sigma = \{q_\alpha, q_\beta\}$, that if $u(v = A|s_i, \sigma) - u(v = B|s_i, \sigma) > 0$ then the probability of voting A increases to $\tilde{q}_\alpha = q_\alpha + \tilde{\varepsilon}$ (with $\tilde{q}_\alpha$ bound by 1), where ε is such that for all $\varepsilon \in (0, \tilde{\varepsilon})$, $\tilde{\sigma} = \{q_\alpha + \varepsilon, q_\beta\} \Rightarrow u(v = A|s_i, \tilde{\sigma}) - u(v = B|s_i, \tilde{\sigma}) > 0$. If the initial utility difference is negative then set $\tilde{q}_\alpha = q_\alpha - \tilde{\varepsilon}$. An equilibrium is then 'stable' if after perturbations of σ the repeated application of best response dynamics returns play to the equilibrium.

Unfortunately, however, it may not be the case that these are the only mixing equilibria as, for some parameter values, Φ may not be quasiconcave or quasiconvex. If additional mixing equilibria do exist then similar arguments would show that at least one of these equilibria must be stable. This possibility will also apply to the equilibria described in Theorem 4.6. As the population grows large, however, this possibility disappears. In Section 4.3.2 I will show that for large voting populations there are no additional mixing equilibria and, therefore, applying the 'stability' refinement eliminates all mixed equilibria.

I turn now to the case of lopsided races in which one candidate begins the contest as a strong favorite.

Theorem 4.6 *Suppose that* $\pi \in (p, 1]$, *voters mix for at most one type of signal, and* $q_\alpha \geq 1 - q_\beta$. *For all integers* n *there exists at least three symmetric equilibria:*

1. $q_\beta = 0, q_\alpha = 1$ *(all voters vote A)*

2. $q_\beta = 1, q_\alpha = 0$ *(all voters vote B)*

3. $q_\beta = 1, q_\alpha = \tilde{q}_\alpha \in (0, 1)$ *(β observers vote B, α observers mix)*

Further, there exists an $\tilde{n}$ *such that for all* $n > \tilde{n}$, $\sigma = \{1, q_\beta\}$ *cannot be an equilibrium if* $q_\beta \in (0, 1]$.

Proof: 1. and 2. As proven in Lemma 4.1.

3. As in Theorem 4.5.

For the final claim I will firstly prove that for large n, $\psi(\beta) > \epsilon > 0$, for some ϵ. As $n \to \infty$, $\psi(s_i) \to \varphi(s_i)[P(A|A) - P(B|A)] - (1 - \varphi(s_i))[P(B|B) - P(A|B)]$. If $q_\alpha = 1$ and $q_\beta \in (0, 1)$ then $P(A|A) > P(B|B)$ and $P(B|A) < P(A|B)$. Therefore $\psi(\beta) > \epsilon > 0$ for large n if $\varphi(\beta) > \frac{1}{2}$, which is true if $\pi \in (p, 1]$.

I will now show that $\Delta(s_i) \to 0$ as $n \to \infty$. Simple calculations show that

$$\frac{\binom{2n+2}{n+1}}{\binom{2n}{n}} = \frac{2(2n+1)}{n+1} \to 4^{-} \text{ as } n \to \infty$$

Also, $\gamma_j(1-\gamma_j) < \frac{1}{4}$ for all $j \in \{A, B\}$as $p > \frac{1}{2}$. This implies that there exists an $\omega \in (0,1)$ such that $\Delta(n+1) < w.\Delta(n)$ (where the argument indicates the population size). This implies that $\Delta(n)$ is bounded above by the converging geometric series $\Delta(1)w^{n-1}$. Because $\Delta(1)w^{n-1} \to 0$ as $n \to \infty$, it must also be that $\Delta(n) \to 0$. Therefore, as $n \to \infty$, $\Delta + k\psi \to k\psi$. As $k\psi(\beta) > 0$ if $k > 0$ this implies that as $n \to \infty$ it must be that $\Phi \neq 0$ and β observers can't mix in equilibrium. ∎

These results are depicted graphically in Figures 4.3a and 4.3b. The inclusion of the voters' desire to vote for the winner has expanded the set of equilibria in this environment also. If $k = 0$ then equilibrium 3 would no longer exist (instead there would exist an equilibrium in which α observers always voted for A and β observers mixed – this is the point in Figure 4.3a where $\Delta = 0$). Equilibrium 3 is similar in intuition to those of Theorem 4.5. It is of the most interest as even though A is considered the strong front runner, the self fulfilling beliefs of voters induces α observers to mix, and in fact vote for B a large proportion of the time. In contrast, β observers vote informatively for the underdog candidate. Significantly, as n expands this becomes the only equilibrium in which any information is aggregated.[23]

To understand why equilibria requiring β observers to mix disappear as the population grows we must look at how the functions Δ and ψ behave as n gets large. These characteristics will be critical to the results of the following section in which I consider asymptotic equilibrium properties. The function Δ is weighted by the voter's pivot probability. As the population increases this probability becomes smaller. Therefore, Δ approaches zero for all values of q_β as n increases. In contrast, ψ remains bounded away from, and above, zero (when α voters always vote for A). Thus, $\Delta + k\psi \to \varepsilon > 0$ as $n \to \infty$ and the mixing equilibria disappear. Essentially, the utility from being in the majority begins to dominate the utility that could arise from being pivotal, and

[23] This result suggests a question about the optimal size of a jury, or voting population. If population increases cause the most efficient equilibria to disappear then second best outcomes may be achieved by restricting the population to a certain size. As I am interested in information aggregation in large populations I will not pursue this question here, though I hope to take it up in future work.

Key: $\Delta \equiv$ ·················· , $\psi \equiv$ ------------ ,

$\Phi = u(v = A) - u(v = B) = \Delta + k\psi \equiv$ ————

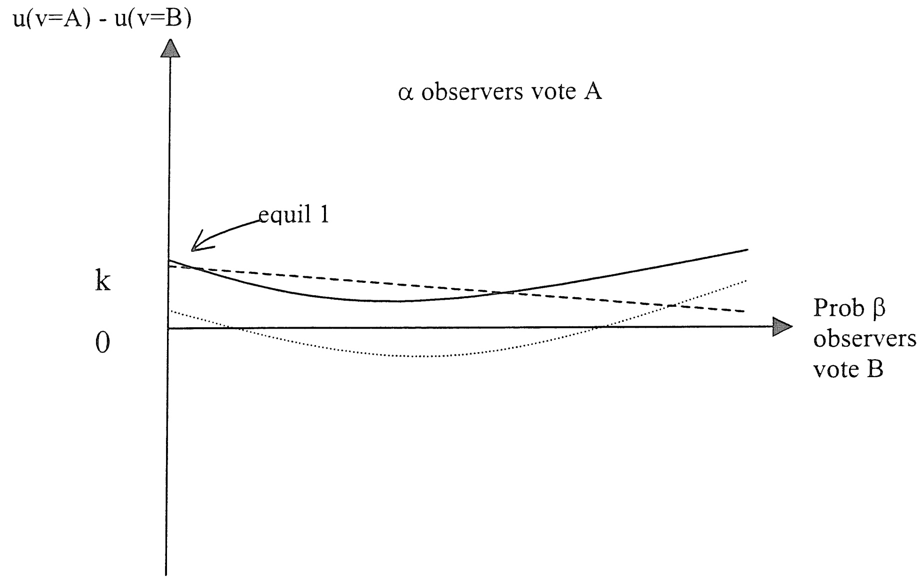


Figure 4.3a: $\pi \in (p,1]$

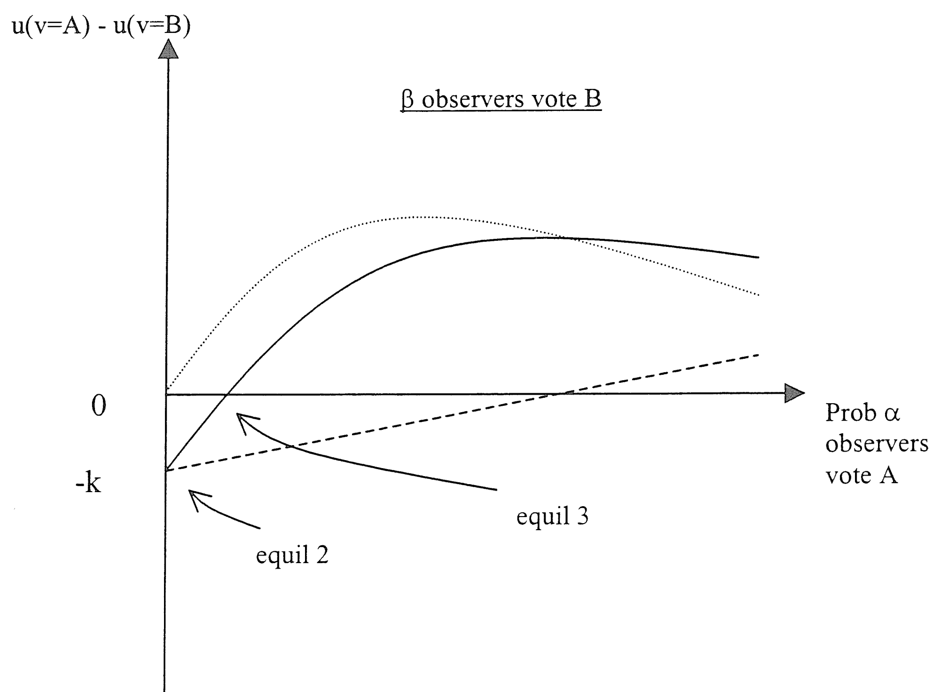


Figure 4.3b: $\pi \in (p,1]$

Figure 4.3: Finite Population Equilibria in Lopsided Elections

in the lopsided race all β observers follow the lead of α observers and support the strong front running candidate.

4.3.2 Optimal Equilibria

In this section I will consider the optimal symmetric equilibria of the voting game as the population becomes large. I will then compare the information aggregation capabilities of these equilibria as the limit is approached to the equilibria at the limit reported in Section 4.2.2. Once again I will consider the equilibria in two cases. Denote by P_n^{sim} the maximum probability that the "better" candidate is selected for population size n.

In large electorates each voters' pivot probability approaches zero and their behavior is dictated primarily by their desire to vote for the winner. Significantly, however, Δ does not completely disappear in finite populations as it does in the limit results of Section 4.2.2, and still affects the set of equilibria even in large populations. The following lemma turns to a consideration of ψ and characterizes its behavior in large electorates.

Lemma 4.4 *Suppose $n \to \infty$, and consider a sequence of strategy profiles $\sigma^n = \left\{q_\alpha^n, q_\beta^n\right\}$.*

1. *If $q_\alpha^n = 1\ \forall n$ then $P(A|A, \sigma^n) \to 1$ and $\psi(\beta|\sigma^n) = 0 \Rightarrow P(B|B, \sigma^n) \to \frac{1}{2(1-\varphi)}$.*
2. *If $q_\beta^n = 1\ \forall n$ then $P(B|B, \sigma^n) \to 1$ and $\psi_n(\beta|\sigma^n) = 0 \Rightarrow P(A|A, \sigma^n) \to \frac{1}{2\varphi}$.*

Proof: $q_\alpha = 1 \Rightarrow \gamma_A > p > \frac{1}{2}$. The law of large numbers therefore implies that $P(A|A, q_\alpha, q_\beta) \to 1$.

ψ can be expressed as the following:

$$\begin{aligned}\psi &= \varphi\left[P(A|A, q_\alpha, q_\beta) - P(B|A, q_\alpha, q_\beta)\right] \\ &\quad + (1-\varphi)\left[1 - P(B|B, q_\alpha, q_\beta) - P(B|B, q_\alpha, q_\beta)\right]\end{aligned}$$

As $n \to \infty$, $\psi \to \varphi[1-0]+(1-\varphi)[1-2P(B|B,q_\alpha,q_\beta)]$. Thus, $\psi_n(q_\beta^n)=0 \Rightarrow \varphi + (1-\varphi)[1-2P(B|B,q_\alpha,q_\beta)]=0$. Rearranging this expression gives $P(B|B,q_\alpha,q_\beta) = \frac{1}{2(1-\varphi)}$. A similar analysis proves Case 2. ∎

There are several things to note from this lemma: firstly, that $\frac{1}{2(1-\varphi)}$ is increasing in φ, and secondly that $\frac{1}{2(1-\varphi)} \in [0,1]$ if and only if $\varphi \in \left[0,\frac{1}{2}\right]$, which requires $\pi \in [0,p]$. Thus for large n, $\psi(\beta)=0$ can only occur if $\pi \le p$, which supports the findings of Theorem 4.6 that equilibria requiring β observers to mix when A is the front runner disappear for large n.

These results are critical to understanding the behavior of the equilibria in large populations. Suppose an equilibrium is being played such that β observers mix and α observers vote A. Then as $n \to \infty$, $P(B|B,q_\alpha,q_\beta) \to \frac{1}{2(1-\varphi)}$ requires $\gamma_\beta \to \frac{1}{2}$. As the limit is approached A wins in state A with probability one, and B wins with probability $\in \left(\frac{1}{2},1\right)$, and therefore some information is aggregated in each state. The equilibrium is achieved because each voter believes when she is pivotal that it is more likely that B is the "better" candidate (thus the condition that $\varphi < \frac{1}{2}$). The fact that $P(B|B,\sigma^n) \notin \left\{\frac{1}{2},1\right\}$ is critical to this result. If B wins with probability one then the voter strictly prefers voting for B (as she still prefers B if pivotal and now B is also more likely to win). Alternatively, if the probability is $\frac{1}{2}$ then the voter strictly prefers to vote for A as she has an equal chance of being in the majority in state B and a strictly greater chance in state A (which dominates the Δ term in large elections).

Significantly, however, it is not possible at the limit for $P(B|B,q_\alpha,q_\beta) \in \left(\frac{1}{2},1\right)$. Therefore, at the limit this equilibrium which aggregates some information no longer exists. This characteristic dictates the relative information aggregation capabilities of the voting mechanism in large finite populations as compared to infinite populations. In the following two theorems I characterize the information aggregation capabilities of a large finite voting population.

Theorem 4.7 *Suppose that* $\pi \in [1-p,p]$. *As* $n \to \infty$ *the optimal equilibria is:* $\sigma = \{1,1\}$ *(all voters vote informatively) and* $P_n^{sim} \to 1$.

Proof: Denote by P_i the probability of selecting the "better" candidate as $n \to \infty$ for equilibrium i from Theorem 4.5. This probability is given by,

$$P_i = \pi P\left(A|A, q_\alpha, q_\beta\right) + (1-\pi)\, P\left(B|B, q_\alpha, q_\beta\right)$$

Obviously, $P_1 = \pi$ and $P_2 = 1 - \pi$. From the law of large numbers, and Lemma 4.4,

$$\begin{aligned} P_3 &= \pi.1 + (1-\pi).1 = 1 \\ P_4 &= \pi.1 + (1-\pi).\frac{1}{2(1-\varphi)} < 1 \\ P_5 &= \pi.\frac{1}{2\varphi} + (1-\pi).1 < 1 \end{aligned}$$

Additionally, $P_4 = 1$ for $\varphi = \frac{1}{2}$, which requires $P\left(B|B, q_\alpha, q_\beta\right) = 1$. For $\gamma_B = 1$ this is identical to equilibrium 3, and for $\gamma_B < 1$ this is dominated by equilibrium 3 (as with greater probability B doesn't win when it is the "better" candidate). Likewise for P_5. ∎

This theorem shows that in tight elections the optimal equilibrium of the voting game in large populations is exactly the same as when the population is infinite. Thus the optimal equilibrium for tight elections of Section 4.2.2 is an appropriate substitute for large finite populations. I now turn to lopsided elections.

Theorem 4.8 *Suppose that $\pi \in (p, 1]$, and $n \to \infty$. If*

1. *$p \geq \frac{1}{\sqrt{2}}$ the uniquely optimal equilibrium is $\sigma = \{1, 0\}$ (all voters vote A) and $P_n^{sim} = \pi$.*

2. *$p \in \left(\frac{1}{2}, \frac{1}{\sqrt{2}}\right)$ and $\pi > \frac{1+p}{1+2p}$ the uniquely optimal equilibrium is $\sigma = \{1, 0\}$ (all voters vote A) and $P_n^{sim} = \pi$.*

3. *$p \in \left(\frac{1}{2}, \frac{1}{\sqrt{2}}\right)$ and $\pi < \frac{1+p}{1+2p}$ the uniquely optimal equilibrium is $\sigma = \{\tilde{q}_\alpha, 1\}$, where $\tilde{q}_\alpha \in (0, 1)$ (β observers vote B, α observers mix) and $P_n^{sim} \to \frac{1}{2}\left(1 + \frac{1-\pi}{p}\right)$.*

4. $p \in \left(\frac{1}{2}, \frac{1}{\sqrt{2}}\right)$ *and* $\pi = \frac{1+p}{1+2p}$ *then the optimal equilibria are* $\sigma = \{1, 0\}$ *(all voters vote A) and* $\sigma = \{\tilde{q}_\alpha, 1\}$, *where* $\tilde{q}_\alpha \in (0, 1)$ *(*β *observers vote B,* α *observers mix). For both equilibria* $P_n^{sim} \to \pi$.

Proof: Using the definitions of Theorem 4.7, $P_1 = \pi, P_2 = 1 - \pi, P_3 = \pi.\frac{1}{2\varphi} + (1-\pi)$. Obviously $P_1 > P_2$, so compare P_1 and P_3. $P_1 > P_3 \Rightarrow \pi > \frac{1+p}{1+2p} = \chi$ as $\varphi(\alpha) = \frac{p\pi}{p\pi + (1-p)(1-\pi)}$ (which proves Claim 2). Note that $\frac{\partial \chi}{\partial p} = \frac{-1}{(1+2p)^2} < 0$, and $\chi = p \Rightarrow p = \frac{1}{\sqrt{2}}$. Therefore, for $p \geq \frac{1}{\sqrt{2}}$ it must be that $P_1 > P_3$ as $\pi > p$ and Claim 1 is proven. Further, $P_1 < P_3 \Rightarrow \pi < \frac{1+p}{1+2p}$, which proves the first part of Claim 3. Noting that $\varphi = \frac{p\pi}{p\pi + (1-p)(1-\pi)}$, and substituting into P_3 gives

$$\begin{aligned} P_3 &= \frac{\pi}{2} \cdot \frac{p\pi + (1-p)(1-\pi)}{p\pi} + (1-\pi) \\ &= \frac{\pi}{2} + (1-\pi)\left[\frac{1-p}{2p} + 1\right] = \frac{1}{2}\left(1 + \frac{1-\pi}{p}\right). \end{aligned}$$

Claim 4 then follows from the equalities of these expressions. ∎

Thus, for large finite populations one of two very different equilibria may be optimal in lopsided races. One of these equilibria does not aggregate information at all and requires all voters to support the front runner. The other equilibrium, which dominates for certain parameter values, requires some informative voting. Strangely, however, this equilibrium requires that informative voting be more likely for the underdog than for the front runner. This leads to the counterintuitive, and somewhat strange, conclusion that the underdog is actually more likely to win the election! In this equilibrium then it is not exactly clear what it means to be the underdog. This equilibrium requires observers of β signals to always vote for B, and for α signal observers to mix. In large populations this implies that B always wins when he is the "better" candidate, but that A only wins a fraction of the time when he is "better."

Relative to the uninformative voting equilibria, this leads to superior outcomes when B is the "better" candidate, but inferior outcomes when A is the "better" candidate. Therefore, the mixing equilibria is preferred when the prior belief is not too lopsided towards A, as then the expected cost is smaller and the expected benefit greater. This implies that for very lopsided races society is better off if the voters do not use their information rather than play the mixing equilibria. Strangely enough, if this equilibrium is to be played then holding the election actually has a negative effect on social welfare (as measured by the probability of selecting the "better" candidate).[24]

It should be noted that the parameter values covered by Claim 1 of Theorem 4.8 are very broad. If a signal is moderately informative ($p \geq 0.707$) then the large but finite population results are identical to the infinite population results. Even for less accurate signals the two sets of results are still identical for a large range of parameters (Claim 2). For example, if $p = \frac{3}{5}$ then the infinite population results correspond exactly with the finite population results if $\pi \geq \frac{8}{11}$. Further, even if the results are not identical, the same substantive conclusions can be drawn: for tight races information aggregation is perfect, but at $\pi = p$ there is a discontinuity and information aggregation is not complete, and approaches zero as priors become more lopsided.

Figures 4.4a and 4.4b capture these relationships graphically, plotting P_n^{sim} for large n against different prior beliefs. The first panel shows the situation described by Claims 1 and 2 of Theorem 4.8, and the second panel shows the situations described by Claim 3 (Claim 4 is the point at which the finite population curve kinks). Both panels also show the findings from Theorem 4.7 (that are unaffected by the cases demarcated in Theorem 4.8). In the limiting case of an infinite population, as presented in Section 4.2.2, the optimal equilibria is for all voters in a lopsided race to vote for the front runner, implying that $P_n^{sim} = \pi$. Therefore, the results for large finite populations and an infinite population are equivalent for the parameters covered by Claims 1 and 2. It is only in Claim 3 that they differ, and then only for certain prior beliefs.

[24]An interesting question then is, if voters are given the option to abstain, does this equilibrium still exist? An initial suspicion is that the answer depends critically upon the exact specification of the utility from voting for the winner. In the current model voters gain utility from voting for the winner and therefore are likely to turn out even if an inefficient equilibria is being played. However, alternative specifications may result in abstention. For example, if voters also incurred a disutility from voting for the loser then an uncertain and inefficient outcome may induce them to abstain.

Figure 4.4a: $p \geq \frac{1}{\sqrt{2}}$ or $p \in (\frac{1}{2}, \frac{1}{\sqrt{2}})$ and $\pi \geq \frac{1+p}{1+2p}$

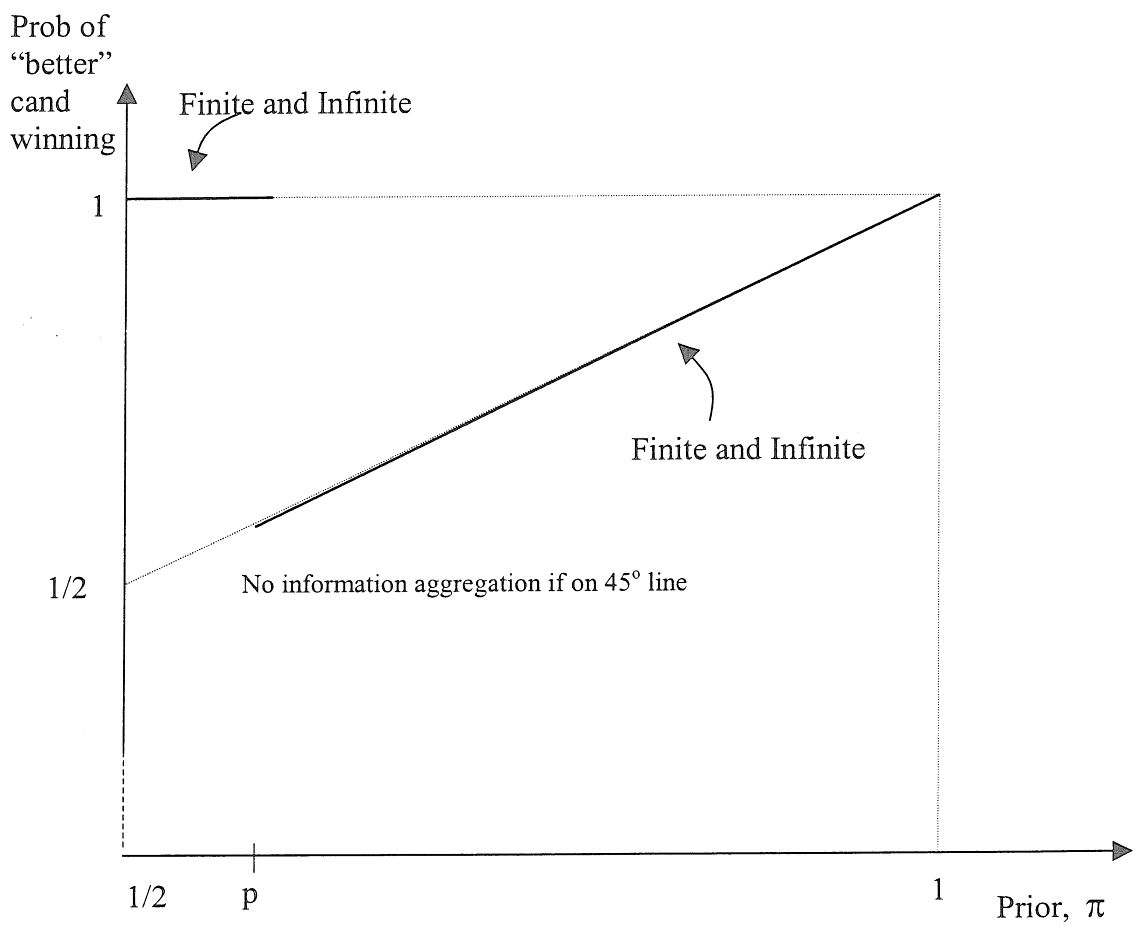
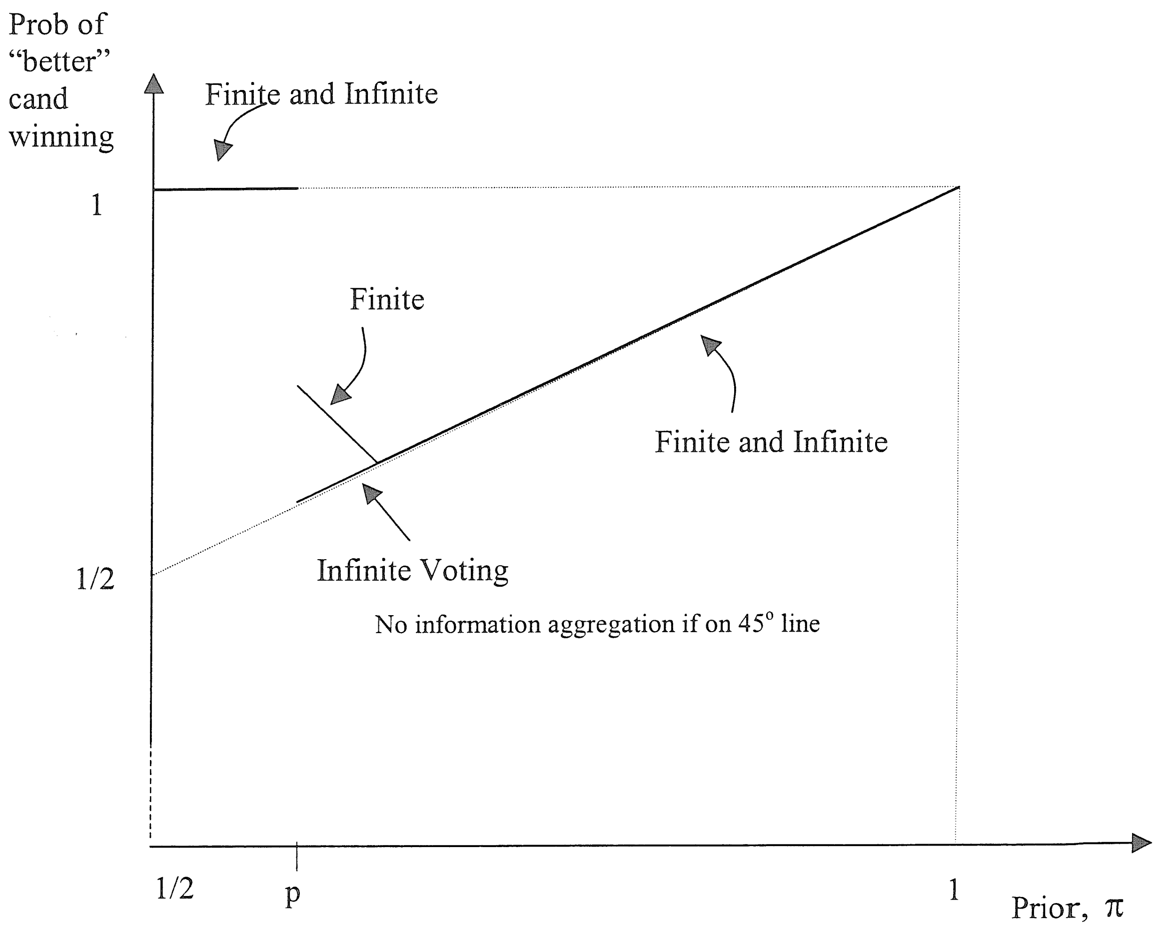


Figure 4.4: Simultaneous Voting: Infinite vs. Finite Populations (Optimal Equilibria)

Figure 4.4b: $p \in (\frac{1}{2}, \frac{1}{\sqrt{2}})$ and $\pi < \frac{1+p}{1+2p}$



In fact, if the stability requirement is applied then the mixing equilibria are eliminated and the set of optimal equilibria for large finite and infinite populations are equivalent everywhere. This result is expressed in Proposition 4.2 below. As equivalence in tight races was established in Theorem 4.7, I will focus here solely on lopsided races. Firstly I require the following lemma that proves there are no equilibria in addition to those described in Theorem 4.6 as the population becomes large.

Lemma 4.5 *Suppose that* $\pi \in (p, 1]$. *Then* $\exists \overline{n}$ *such that* $\forall n > \overline{n}$ *there is a unique* $\tilde{q}_\alpha \in (0,1)$ *where* $q_\beta = 1, q_\alpha = \tilde{q}_\alpha$ *is an equilibrium.*

Proof: For this to be the case there must be only one $q_\alpha \in (0,1)$ such that $\Phi(\alpha) = 0$ (for exposition I will drop the α argument and reinstall the strategy argument σ, as determined by q_α). To establish this result, I consider the components of $\Phi(q_\alpha) = \Delta(q_\alpha) + k\psi(q_\alpha)$ separately (see page 186 for this specification). In Theorem 4.6 it was shown that $\forall q_\alpha$, $\Delta(q_\alpha) \to 0$ as $n \to \infty$. Now consider the following decomposition, $\psi = \psi^1 - \psi^2$, where $\psi^1(q_\alpha) = \varphi \sum_{j=0}^{n-1} \binom{2n}{j} \left[\gamma_A^{2n-j}(1-\gamma_A)^j - \gamma_A^j (1-\gamma_A)^{2n-j}\right]$, and $\psi^2(q_\alpha) = (1-\varphi)\sum_{j=0}^{n-1}\binom{2n}{j}\left[\gamma_B^{2n-j}(1-\gamma_B)^j - \gamma_B^j(1-\gamma_B)^{2n-j}\right]$. As $q_\beta = 1$ it must be that $\gamma_B > \frac{1}{2}$. Therefore, by the law of large numbers, $\psi^2(q_\alpha) \to (1-\varphi)\ \forall q_\alpha$ as $n \to \infty$ (this can most easily be seen by recalling that the components of ψ^2 are probabilities). Similarly it can be seen that $\psi^1(0) = -\varphi, \psi^1(1) = \varphi$, and that $\frac{\partial \psi^1(q_\alpha)}{\partial q_\alpha} > 0\ \forall q_\alpha \in (0,1)$. Therefore, combining these facts, if n is large then for any $\tilde{q}_\alpha$ such that $\Phi(\tilde{q}_\alpha) = 0$ there must be an $\varepsilon > 0$ such that $\tilde{q}_\alpha = \overline{q}_\alpha + \varepsilon$ where $\overline{q}_\alpha$ is the unique solution to $\psi^1(q_\alpha) = 1 - \varphi$ (this point must exist as $\varphi(\alpha) > \frac{1}{2}$). Further, and most critically, if there are two such roots of $\Phi(q_\alpha) = 0$, denoted q' and q'', then $\psi^1(q') - \psi^1(q'')$ must be arbitrarily close to zero. I will now establish a contradiction with respect to this requirement.

Consider the behavior of ψ^1 as $n \to \infty$, denoting the population size by a subscript. As was the case with $\psi^2(q_\alpha)$ for a given q_α as $n \to \infty$, $\gamma_A > \frac{1}{2} \Rightarrow \psi_n^1(q_\alpha) \to \varphi$, $\gamma_A < \frac{1}{2} \Rightarrow \psi_n^1(q_\alpha) \to -\varphi$, and if $\gamma_A = \frac{1}{2} \Rightarrow \psi_n^1(q_\alpha) = 0$. I will now establish the rates of convergence to these limits. For $n = 1$, $\psi_1^1(q_\alpha) = \varphi(2\gamma_A - 1)$, which is linear. Thus, for $q'' > q'$, $\psi_1^1(q'') - \psi_1^1(q') > \rho > 0$. Further, the following relationship for ψ_n^1

as it varies in n must hold,

$$\begin{aligned}\psi^1_{n+1}(q_\alpha) &= \psi^1_n(q_\alpha) - \left[\binom{2n}{n-1}\gamma_A^{n+1}(1-\gamma_A)^{n-1}\right](1-\gamma_A)^2 \\ &\quad + \left[\binom{2n}{n}\gamma_A^n(1-\gamma_A)^n\right]\gamma_A^2 + \left[\binom{2n}{n-1}\gamma_A^{n-1}(1-\gamma_A)^{n+1}\right]\gamma_A^2 \\ &\quad - \left[\binom{2n}{n}\gamma_A^n(1-\gamma_A)^n\right](1-\gamma_A)^2 \\ &= \psi^1_n(q_\alpha) + \binom{2n}{n}\gamma_A^n(1-\gamma_A)^n(2\gamma_A - 1)\end{aligned}$$

Therefore, for $\gamma_A > \frac{1}{2}$, $\psi^1_{n+1}(q_\alpha) - \psi^1_n(q_\alpha) > 0$ as $(2\gamma_A - 1) > 0$. Further, $\frac{\partial}{\partial\gamma_A}\left[\frac{\partial[\psi^1_{n+1}(q_\alpha)-\psi^1_n(q_\alpha)]}{\partial n}\right] < 0$. Suppose for some finite $\overline{n}$ that $\psi^1_{\overline{n}}(q') - \psi^1_{\overline{n}}(q'')$ is arbitrarily small. Therefore, $\overline{n}$ must be large enough such that $\psi^1_{\overline{n}+1}(q') - \psi^1_{\overline{n}}(q') > \psi^1_{\overline{n}+1}(q'') - \psi^1_{\overline{n}}(q'')$. As roots of the equation $\Phi(q_\alpha) = 0$ must occur for values of q_α such that $\psi^1_n(q_\alpha)$ is in some neighborhood of $(1-\varphi)$, it must be that $\sum_{n=\overline{n}+1}^{\infty}\left[\psi^1_{n+1}(q') - \psi^1_n(q')\right] > 0$. This implies that $\exists\delta > 0$ such that $\sum_{n=\overline{n}+1}^{\infty}\left[\psi^1_{n+1}(q') - \psi^1_n(q')\right] + \delta > \sum_{n=\overline{n}+1}^{\infty}\left[\psi^1_{n+1}(q'') - \psi^1_n(q'')\right]$. Consequently, as $n \to \infty$, it must be that $\psi^1_n(q') \nrightarrow \psi^1_n(q'')$, which contradicts the fact that $\psi^1_n(q')$, $\psi^1_n(q'') \to \varphi$. Thus, there can only be one solution to $\Phi(q_\alpha) = 0$ and the mixed equilibrium is unique. ∎

With the uniqueness of the mixed strategy equilibrium for large populations established, I can now state complete equivalence of infinite and large finite voting populations under the 'stability' refinement. To establish this result I need only prove that the unique mixed strategy equilibrium in lopsided races, as shown by the Lemma above, is unstable with respect to small perturbations and best response dynamics. To establish this result I need only prove that the unique mixed strategy equilibrium in lopsided races, as shown by the lemma above, is unstable with respect to small perturbations and best response dynamics.[25]

Proposition 4.2 *For all $\pi \in [0,1]$ the optimal 'stable' equilibrium is the same for large finite and infinite voting populations.*

[25]See footnote 22 on page 192 for the definition of 'stability' employed here.

Proof: At the equilibrium, $\sigma = \{\tilde{q}_A, 1\}$, $u(v = A|\beta, \sigma) - u(v = B|\beta, \sigma) > 0$ and $u(v = A|\alpha, \sigma) - u(v = B|\alpha, \sigma) = 0$. Further, for $\varepsilon > 0$, $\sigma' = \{\tilde{q}_A + \varepsilon, 1\} \Rightarrow u(v = A|\alpha, \sigma') - u(v = B|\alpha, \sigma') > 0$ and $\sigma'' = \{\tilde{q}_A - \varepsilon, 1\} \Rightarrow u(v = A|\alpha, \sigma'') - u(v = B|\alpha, \sigma'') < 0$. Therefore, consider the perturbation in favor of A, $\sigma^* = \{\tilde{q} + \Delta, 1 - \Delta\}$ for small Δ. Best response dynamics results in the strategy $\sigma^{**} = \{\tilde{q} + \Delta + \kappa, 1\}$ for some $\kappa > 0$. Therefore, play does not converge back to σ, and the equilibrium is not stable. ∎

Thus, it would seem that the infinite population results of Section 4.2.2 are not an inappropriate proxy for the equilibria that could be expected in large finite populations and, in fact, under the 'stability' refinement the optimal equilibria are completely equivalent.

4.4 Conclusion

In this chapter I have compared the information aggregation properties of the sequential and simultaneous voting processes. I showed that they are generally not equivalent once the utility function of voters is extended in a natural way to include a desire to vote for the winning candidate. Indeed, I find that it is only in the limiting case when this desire is zero that the rules are equivalent. Further, I characterize the conditions when each rule is preferred (i.e., selects the "better" candidate with greater probability).

I find that in close races sequential voting does give disproportionate influence to early voters and consequently simultaneous voting is preferred. In contrast, for lopsided races this power still exists but in this context it proves beneficial. Bandwagon voting implies that the pivot probabilities of early voters don't go to zero, even as the size of the population goes to the limit, thus ensuring some informative voting. When voting is simultaneous all pivot probabilities go to zero and the incentive to be in the winning majority dominates, precluding any informative voting. This leads to the surprising result that in lopsided races sequential voting is preferred precisely because bandwagons occur. These results are consistent with existing suspicions and observations of the effects of front-loading in U.S. primary elections.

I will conclude the chapter with several comments on assumptions made herein and directions for future research. In Section 4.2.3 simultaneous voting is associated with a front-loaded primary. Obviously, the front loading of primaries is not complete and the actual structure is somewhere between a simultaneous and a completely sequential election. Thus, to analyze the effects of front-loading completely, and search for the optimal primary structure, the model would need to be generalized to allow for purely sequential voting, purely simultaneous voting, and any hybrid in between. I intend to consider this in future work.[26]

Further differences in the information aggregation properties of the two processes may arise if the model is extended to permit competition amongst multiple candidates. A well known result in the simultaneous voting literature is that in all stable equilibria of a multiple candidate voting game only two candidates will receive votes (see Fey (1997)). However, in the sequential voting game it is possible that the logic of the two candidate model carries over, and more than two candidates will receive votes (at least early in the sequence). There may exist an equilibrium with multiple thresholds that the beliefs about the trailing candidates must exceed for these candidates to receive informative votes and stay in the race. This capability, if true, would alter considerably the information aggregation properties of the sequential voting process relative to simultaneous voting and, therefore, warrants further attention.

The model assumed that under both sequential and simultaneous voting each voter receives an independent and identically distributed private signal. Such an assumption would seem appropriate in a sequential voting game as in the presidential primaries voters mostly learn only "horse race" information until their state's primary.[27] However, in a simultaneous election the assumption of independence may not be so appropriate. It is not clear whether such correlation would improve or

[26] Such a general voting structure was considered by Dekel and Piccione (2000). In their model they found that the optimal equilibrium was unaffected by the actual timing of votes (when voting strategies are monotonic). However, as can already be seen, this result will not hold when voters have a desire to vote for the winning candidate.

[27] Popkin (1994).

harm the information aggregation properties of this process.

When considering the application of these results to U.S. presidential primaries and the onset of front-loading it becomes clear that the timing of votes should actually be endogenous to the model. That is, the states themselves are able to choose the timing of their vote in order to further their self interest. Thus, there may exist a tension between what vote order is the best for the parties themselves or society as a whole, and what arises from an equilibrium of the game amongst states to select their vote position. This potential inefficiency could have serious implications for the selection of American leaders. I also intend to look further at this possibility.

Bibliography

[1] Aldrich, J.H. (1980), *Before the Convention: Strategies and Choices in Presidential Nomination Campaigns* (Chicago, Illinois: The University of Chicago Press).

[2] Alesina, A. and H. Rosenthal (1995), *Partisan Politics, Divided Government, and the Economy* (Cambridge, U.K.: Cambridge University Press).

[3] Alvarez, R.M. and G. Glasgow (1997), "Do Voters Learn from Presidential Election Campaigns?" Working Paper #1022, California Institute of Technology.

[4] Alvarez, R.M. and J. Nagler (1999), "A New Approach for Modeling Strategic Voting in Multiparty Elections," *British Journal of Political Science* forthcoming.

[5] Arian, A. (1998), *The Second Republic: Politics in Israel* (Chatham, NJ: Chatham House Publishers).

[6] Aronson, E., T.D. Wilson, and R.M. Akert (1997), *Social Psychology* (New York, NY: Longman).

[7] Austen-Smith, D. (1981), "Party policy and campaign costs in a multi-constituency model of electoral competition," *Public Choice* 37: 389-402.

[8] Austen-Smith, D. and J. Banks (1988), "Elections, Coalitions, and Legislative Outcomes," *American Political Science Review* 82: 409-422.

[9] Austen-Smith, D. and J. Banks (1996), "Information Aggregation, Rationality, and the Condorcet Jury Theorem," *American Political Science Review* 90: 34-45.

[10] Banerjee, A.V. (1992), "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics* 107: 797-817.

[11] Banks, J. and J. Sobel (1987), "Equilibrium Selection in Signalling Games," *Econometrica* 55: 647-662.

[12] Bartels, L.M. (1985), "Expectations and Preferences in Presidential Nominating Campaigns," *American Political Science Review* 79: 804-815.

[13] Bartels, L.M. (1988), *Presidential Primaries and the Dynamics of Public Choice* (Princeton, New Jersey; Princeton University Press).

[14] Berelson, B.R., P.F. Lazarsfeld, and W.N. McPhee (1954), *Voting* (Chicago, Illinois: The University of Chicago Press).

[15] Besley, T. and S. Coate (1997), "An Economic Model of Representative Democracy," *The Quarterly Journal of Economics* 112: 85-114.

[16] Bikhchandani, S., D. Hirshleifer and I. Welch (1992), "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy* 100: 992-1026.

[17] Brams, S. and P. Fishburn (1983), *Approval Voting* (Boston, MA; Brikhauser).

[18] Brownstein, R. (2000), "It's Now, if Ever: 2 Challengers Pin Hopes on New Hampshire," *Los Angeles Times* January 26: A8.

[19] Cain, B., J. Ferejohn, and M. Fiorina (1987), *The Personal Vote: Constituency Service and Electoral Independence* (Cambridge, MA: Harvard University Press).

[20] Callander, S. (1999), "Electoral Competition with Entry," Working Paper #1083, California Institute of Technology.

[21] Calvert, R. (1985), "Robustness of the Multidimensional Voting Model: Candidates' Motivations, Uncertainty and Convergence," *American Journal of Political Science* 79: 1056-1070.

[22] Canon, B.C. (1978), "Factionalism in the South: A Test of Theory and a Revisitation of V.O. Key," *American Journal of Political Science* 22: 833-848.

[23] Cherry, T.L. and S. Kroll (2000), "Sincere versus Strategic Voting in Primaries: An Experimental Investigation," Working Paper.

[24] Cho, I.K. and D.M. Kreps (1987), "Signalling Games and Stable Equilibria," *Quarterly Journal of Economics* 102: 179-221.

[25] Cox, G. (1985), "Electoral Equilibrium Under Approval Voting," *American Journal of Political Science* 29: 112-118.

[26] Cox, G. (1987), "Electoral Equilibrium Under Alternative Voting Institutions," *American Journal of Political Science* 31: 82-108.

[27] Cox, G. (1990), "Multicandidate Spatial Competition," in *Advances in the Spatial Theory of Voting*, eds. J. Enelow and M. Hinich (Cambridge, U.K.: Cambridge University Press).

[28] Cox, G. (1997), *Making Votes Count* (Cambridge, U.K.: Cambridge University Press).

[29] Crewe, I. and A. King (1994), "Are British Elections Becoming More 'Presidential,"' in *Elections at Home and Abroad*, eds. M.K. Jennings and T.E. Mann (Ann Arbor, MI; The University of Michigan Press).

[30] Dekel, E. and M. Piccione (2000), "Sequential Voting Procedures in Symmetric Binary Elections," *Journal of Political Economy* 108: 34-55.

[31] Downs, A. (1957), *An Economic Theory of Democracy* (New York, New York: Harper and Row).

[32] Duverger, M. (1954), *Political Parties* (New York, NY: Wiley).

[33] Feddersen, T., I. Sened, and S. Wright (1990), "Rational Voting and Candidate Entry under Plurality Rule," *American Journal of Political Science* 34: 1005-16.

[34] Feddersen, T. and W. Pesendorfer (1997), "Voting Behavior and Information Aggregation in Elections with Private Information," *Econometrica* 65: 1029-58.

[35] Feddersen, T. and W. Pesendorfer (1998), "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting," *American Political Science Review* 92: 23-35.

[36] Fey, M. (1997), "Stability and Coordination in Duverger's Law: A Formal Model of Preelection Polls and Strategic Voting," *American Political Science Review* 91: 135-147.

[37] Fey, M. (1998), "Informational Cascades and Sequential Voting," Working Paper, University of Rochester.

[38] Fiore, F. (2000), "N.H. Voters Prepare to Anoint," *Los Angeles Times* January 26: A1.

[39] Greenfield, J. (1982), *The Real Campaign: How the Media Missed the Story of the 1980 Campaign* (New York, NY: Summit Books).

[40] Guarnaschelli, S., R.D. McKelvey, and T.R. Palfrey (2000), "An Experimental Study of Jury Decision Rules," *American Political Science Review* 94: 407-424.

[41] Herron, M.C. (1998), "Voting, Abstention, and Individual Expectations in the 1992 Presidential Election," Working Paper, Northwestern University.

[42] Hinich, M.J. and P.C. Ordeshook (1974), "The Electoral College: A Spatial Analysis," *Political Methodology* Summer: 1-29.

[43] Holcombe, A. (1910), "Direct Primaries and the Second Ballot," *The American Political Science Review* 5: 535-552.

[44] Hotelling, H. (1929), "Stability in Competition," *The Economic Journal* 39: 41-57.

[45] Hung, A. and C.R. Plott (2000), "Information Cascades: Replication and an Extension to Majority Rule and Conformity Rewarding Institutions," *American Economic Review* forthcoming.

[46] Lee, I.H. (1993), "On the Convergence of Informational Cascades," *Journal of Economic Theory* 61: 395-411.

[47] Lijphart, A. (1994), *Electoral Systems and Party Systems* (Oxford, U.K.: Oxford University Press).

[48] Lohmann, S. (1994), "Information Aggregation Through Costly Political Action," *American Economic Review* 84: 518-530.

[49] Lowell, A.L. (1896), *Government and Politics of Continental Europe* (Boston, MA; Houghton Mifflin).

[50] Morton, R.B. and K.C. Williams (2000), *Learning by Voting: Sequential Choices in Presidential Primaries and Other Elections* (Ann Arbor, Michigan: University of Michigan Press).

[51] Myerson, R.B. (1993), "Incentives to Cultivate Favored Minorities Under Alternative Electoral Systems," *The American Political Science Review* 87: 856-869.

[52] Myerson, R.B. and R.J. Weber (1993), "A Theory of Voting Equilibria," *The American Political Science Review* 87: 102-114.

[53] Niemi, R.G. and L. Bartels (1984), "The Responsiveness of Approval Voting to Political Circumstances," *Political Science and Politics* 17: 571-77.

[54] Osborne M.J. (1993), "Candidate Positioning and Entry in a Political Competition," *Games and Economic Behavior* 5: 133-151.

[55] Osborne, M.J. (1995), "Spatial Models of Political Competition Under Plurality Rule: A Survey of Some Explanations of the Number of Candidates and the Positions They Take," *Canadian Journal of Economics* May: 261-301.

[56] Osborne, M.J. and A. Slivinski (1996), "A Model of Political Competition with Citizen-Candidates," *Quarterly Journal of Economics* 111: 65-96.

[57] Osborne, M.J. (2000), "Entry-Deterring Policy Differentiation by Electoral Candidates," forthcoming, Mathematical Social Sciences.

[58] Palfrey, T. (1984), "Spatial Equilibrium with Entry," *Review of Economic Studies* 51: 139-156.

[59] Palfrey, T. (1989), "A Mathematical Proof of Duverger's Law," in *Models of Strategic Choice in Politics*, ed. P. Ordeshook (Ann Arbor, MI: University of Michigan Press).

[60] Popkin, S.L. (1994), *The Reasoning Voter* (Chicago, Illinois: The University of Chicago Press).

[61] Reeves, R. (1977), *Convention* (New York, NY: Harcourt, Brace, Jovanovich).

[62] Riker, W.H. (1982), "The Two-party System and Duverger's Law: An essay on the History of Political Science," *The American Political Science Review* 76: 753-766.

[63] Robertson, D. (1976), *A Theory of Party Competition* (London, U.K.: Wiley).

[64] Schlesinger, J.A. and M. Schlesinger (1990), "The Reaffirmation of a Multiparty System in France," *The American Political Science Review* 84: 1077-1101.

[65] Schneider, W. (1997) "And Now the GOP is Rewriting its Rules," *National Journal* April 12: 734.

[66] Shepsle, K.A. and R.N. Cohen (1990), "Multiparty Competition, Entry, and Entry Deterrence in Spatial Models of Elections," in *Advances in the Spatial Theory of Voting*, eds. J. Enelow and M. Hinich (Cambridge, U.K.: Cambridge University Press).

[67] Shugart, M.S. and R. Taagepera (1994), "Plurality Versus Majority Election of Presidents," *Comparative Political Studies* 27: 323-348.

[68] Shvetsova, O. (1995), "Design of Political Institutions in Divided Societies," Ph.D. dissertation, California Institute of Technology.

[69] Smallwood, F. (1983), *The Other Candidates: Third Parties in Presidential Elections* (Hanover, NH: University Press of New England).

[70] Wit, J. (1997), "Herding Behavior in a Roll-Call Voting Game," Working Paper, CREED, Department of Economics, University of Amsterdam.

[71] Wittman, D. (1983), "Candidate Motivation: A Synthesis of Alternative Theories," *American Political Science Review* 77: 142-157.

[72] Wright, J.F.H. (1986), "Australian Experience with Majority-Preferential and Quota-Preferential Systems," in *Electoral Laws and their Political Consequences*, eds. Grofman and Lijphart (New York, NY: Agathon Press).

[73] Wright, S.G. and W.H. Riker (1989), "Plurality and Runoff Systems and Numbers of Candidates," *Public Choice* 60: 155-175.