ON THE ROLE OF IMPERFECT INFORMATION IN ELECTORAL POLITICS

Thesis by

Randall Lee Calvert

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1980

(Submitted August 7, 1979)

ACKNOWLEDGEMENT

Financial support for the research and writing of this thesis was provided by the Earle C. Anthony Fellowship; the Anna and James McDonnell Memorial Scholarship; a National Science Foundation Graduate Fellowship; and Graduate Research and Teaching Assistantships awarded by the California Institute of Technology.

For their generous assistance, advice, and encouragement, I am grateful to Bruce Cain, John Ferejohn, Morris Fiorina, and Louis Wilde. I would also like to thank George Fox, Gary Miller, and Thomas Palfrey, with whom I have had many helpful discussions about the work presented here. The secretarial assistance of Mary Doan and Dee Scriven and the secretarial and administrative efforts of Susan Davis have been most valuable. Finally, I would like to thank my wife, Betty Ann June, for her constant support and for putting up with all this.

ABSTRACT

Two aspects of the electoral process are modelled formally: rational voting when information is costly and imperfect; and candidate strategies when election outcomes cannot be predicted with certainty, under various assumptions about candidates' goals. In both cases, the problem is to rigorously define and characterize rational behavior when uncertainty is taken explicitly into account.

Voters are seen as engaging in a process of optimal sequential sampling, gathering information and updating prior beliefs about candidate positions while deciding at each stage whether to continue gathering information or to stop and choose a candidate. Under quite general conditions, this process is shown to yield a well-defined rule for rational decisionmaking, characterized by a functional equation. The following properties hold for this process: (1) if the cost of information increases, the set of prior belief states at which further sampling occurs is made smaller; (2) certain kinds of increased uncertainty about the desirability of candidates will increase the value of sampling; (3) under special conditions, the candidate chosen will be the one seen as preferred in the final observation, but this relation may fail under more general assumptions. When voters are assumed to observe only the utility level of candidate platforms, and not the platforms themselves, the conditions of (3) can be generalized. Also, a model of voting and the development of party

identification can be defined, bearing a close resemblance to certain non-rational-choice models, which parsimoniously predicts many observed properties of voting.

Candidates are modelled as either seeking to win office or as seeking to implement preferred policies; and as being either certain or uncertain about the outcome of the election given both candidates' strategy choices. The "median voter" or convergent platforms result of spatial modelling holds whenever candidates can predict the outcome with certainty, or when uncertain candidates seek only to win office. But policy-oriented candidates under uncertainty will never adopt identical platforms in equilibrium.

TABLE OF CONTENTS

Chapter 1

VOTING AND INFORMATION: AN INTRODUCTION AND SURVEY

Can a democratic government be responsive to the needs of its citizens? This question has occupied political philosophers since antiquity, and empirical researchers since the beginning of their enterprise somewhat more recently. Among the latter, students of electoral politics have tended to focus on the voters' side of the electoral process, on the issue of whether citizens can effectively transmit their wishes through voting. One school of thought has viewed the question as one of sociological forces: do the characteristics and interests of social classes (see for example Lipset [1960]), or the features of political culture (Almond and Verba [1963]), or the shared attitudes and role definitions of overlapping primary groups (Berelson et al. [1954]) promote a political system in which democratic values flourish? Most scholars of this bent (especially including those mentioned above) have realized that these sociological forces can be fickle, that it is hard to predict how democratic government may continue or fail in the face of presumably static social characteristics. Certainly voters whose choices are determined by their location among social groups, as are those of Berelson et al. [1954], will have a difficult time responding dynamically to crises in the political environment.

For other scholars, the emphasis indeed came to rest on the individual voter. Do voters respond to political realities in terms of their needs or even the needs of society as a whole; are voters rational? The American Voter set forth three conditions to be met by voters if they are to respond to issues.

1. The issue must be cognized in some form.
2. It must arouse some minimal intensity of feeling.
3. It must be accompanied by some perception that one party represents the person's own position better than do the other parties. (Campbell et al. [1960], pp. 169-170.)

The widespread failure of these conditions was interpreted to mean that voters do not base their decisions on real political issues to the extent required for true democracy (Campbell et al. [1960]; Converse [1966]; Stokes and Miller [1966]). Other scholars have disputed either the extent of voter ignorance (Key [1966]; Pomper [1972]; Boyd [1972]) or the way in which Campbell et al. attempted to operationalize issue voting (RePass [1971]). That the disagreement has not yet been resolved is demonstrated by Margolis [1977].

In The Semisovereign People, Schattschneider [1960] presented a scathing criticism of the Michigan school's (and others') approach to voter rationality and democratic government. His point was that full knowledge about the issues of government is impossible even for the governors themselves, and that it is only sensible (rational, we might say) for the ordinary citizen to operate with severely limited information on the subject. To Schattschneider, the question of how a democracy should work was a question of how it should work given sensible, ordinary citizens with such limited information. The like of these passages from Schattschneider has not been seen since, Key's ([1966], p. 7) assertion that "voters are not fools" notwithstanding:

> Definitions of democracy since the time of Aristotle have been made on the assumption that the "many" in a democracy do the same thing that the "one" does in a monarchy ... But obviously the shift from the "one" to the "many" is not merely a change in the number of people participating in power but a change in the way the power is exercised (Schattschneider [1960], pp. 140-141; emphasis in original).
>
> There is no escape from the problem of ignorance, because nobody knows enough to run the government. Presidents, senators, governors, judges, professors, doctors of philosophy and the like are only a little less ignorant than the rest of us (p. 136; emphasis in original).
>
> Only a pedagogue would suppose that the people must pass some kind of examination to qualify for participation in a democracy (p. 135).
>
> The public is far too sensible to attempt to play the preposterous role assigned to it by the theorists (p. 134).

Schattschneider did not attempt to say whether democracy works; his contribution was to suggest <u>how</u> it would work if it did.

With Schattschneider's ideas in mind, let us return to the subject of the individual voter. We have already noted the problems of sociological approaches in trying to provide a theory of democratic government based on a self-equilibrating social system; basing government responsiveness on magnanimous political leadership is an even less likely approach. It seems that the citizens must act to transmit their wishes to the government and influence the government to respond, and this must be accomplished through voting. So the question of democratic government, from the persepctive of individual behavior, becomes: given the realities of ignorance and uncertainty, do voters rationally transmit their wishes and control their government?

The question of rational voting under imperfect information is not just one to be answered empirically. To begin with, the empirical researcher must know what to look for, because rationality may be hard to recognize when voters must contend with such uncertainty. It is first necessary to characterize, theoretically, how a rational voter should be expected to behave; needless to say, this is a complicated issue, and many approaches may be tried. In this study, we will view the voter as having two kinds of tasks, in keeping with Schattschneider's conception:

to participate (here, to vote) rationally, and to rationally gather information about voting. A less tangible theoretical question must precede the task of characterizing rational behavior, however. It is conceivable that there is no consistent, rational course of action available under which the voter may transmit his wishes to the government. The voting decision is affected by the voter's efforts to gather information and the use he makes of it; at the same time, the citizen's opportunity to vote and the choice offered him affect the way he gathers information. The problem is whether an optimizing choice can be made given such interrelations. If it cannot, then rational voting cannot be defined and citizen control of democratic government may be a logical impossibility.

In this study we will use a model of voting when information is costly and incomplete which captures much of this complexity. We will show how rational voting may take place and begin to characterize it. Finally, we will examine some implications of such informational conditions for election outcomes and hence for the polity as a whole. This undertaking is not without intellectual foundations, so we shall first consider some of these.

1. THE IRRATIONAL VOTER AND INFORMATION

Even scholars who have viewed voters as being essentially irrational, either by hypothesis or empirically, have had sophisticated, explicit ideas about the voters' use of political information. In this section we will examine these ideas as they appear in the work of the Columbia and Michigan schools of voting behavior.

The Columbia school is characterized by a particular sociological viewpoint of electoral politics. Among those scholars, although membership in social groups is said to determine the choice made by each voter, campaign information has an important function as well:

> Campaign propaganda has something like the effect of the [photographic] developer and the pencil shading [of a coin beneath a paper]. It brings the voter's predispositions to the level of visibility and expression. It transforms the latent political tendency into a manifest vote (Lazarsfeld et al. [1948], p. 75).

This transformation takes place because of the "selective attention" of each voter to information consistent with his predisposition (Lazarsfeld et al. [1948], p. 76), a subconscious psychological predisposition (p. 81). These predispositions are operationalized as the "index of political predispositions," based on residence (urban or rural), religion, and socioeconomic status. But much of the individual's information comes from his conversations with acquaintances, who tend to be of the same residence, religion, and socioeconomic status.

> [T]hese people might be said to be learning from each other ... probably taking over the judgments of respected intimates as better informed than their own; still others are no doubt stimulating one another toward freshly emergent viewpoints; certainly many are calling each other's attention to common interests ... (Berelson et al. [1954], p. 122).

The famous "two-step flow of communications" from media to opinion leaders to less active citizens (Lazarsfeld et al. [1948], p. 151) is also part of this process. In the end the voter's opinions reflect objective information but are "distorted in accordance with the subject's predispositions" (Berelson et al. [1954], p. 216). The Columbia school's voter, then, begins with preconceived notions about the political campaign; gathers some easily attainable information, often from trusted sources; and combines these to form his later opinions. This process does not differ radically from that of the statistical decisionmaker, who obtains costly information and updates his prior beliefs.

In contrast to the Columbia school, the scholars of the Michigan school are more social-psychological in their orientation. While in principle their view of social predetermination resembles that of the Columbia school, some writings emphasize more the importance of real political information:

> Perceptions are not free-floating creations of the individual voter or of the small social groupings in which they are shared. ... The flow of historical reality has enormous influence on the electorate's perceptions of its political environment.
>
> ... Yet percept and reality are not the same, and ... we will ultimately have to consider not only the "real" properties of these objects but certain processes of individual psychology as well (Campbell et al. [1960], p. 43).

A part of this connection between historical reality and perception is achieved through the association of political parties with various types of policies or outcomes; Campbell et al. cite the example of the Republicans on farm policy in the 1950s, a connection not only shared by many voters but which <u>changed</u> among voters as Republican farm policy itself changed ([1960], p. 48). In general,

> the role of party as a carrier of attitude that rose in public response to things past may be one of profound significance. To a great extent the image of the Republican and Democratic Parties in 1952 and 1956 was the public's response to issues and events of the past generation, whereas popular perceptions of Eisenhower and Stevenson seemed to be fashioned of more current materials ([1960], p. 60).

Another psychological connection between perceptions and reality is provided by the striving for "cognitive balance," in which "the individual strives to give order and coherence to his image of these [political] objects" ([1960], p. 59), including parties, candidates and issues. Party, being a particularly visible fact, is perhaps foremost among these objects as

> ... a supplier of cues by which the individual may evaluate the elements of politics. The fact that most elements of national politics are far removed from the world of the common citizen forces the individual to depend upon sources of information from which he may learn indirectly what he cannot know as a matter of direct experience. Moreover, the complexities of politics and government increase the importance of having relatively simple cues to evaluate what cannot be matters of personal knowledge ([1960], p. 128).

Somehow, though (perhaps through the magic of joint authorship), this use of party as an informational cue is described in the same study as a grossly irrational form of voting behavior. Because individuals respond to the question of whether they consider themselves Republicans, Democrats, or what, with answers which persist over time, party is said to be unresponsive to short term forces and hence to relevant political issues. "Generally, this tie is a psychological identification" ([1960], p. 121), a concept derived from the socio-psychological notion of identification with a primary social group. The vote becomes simply a matter of adherence to group norms, and instead of acting as an informational cue, "identification with a party raises a perceptual screen through which the individual tends to see what is favorable to his partisan orientation" ([1960], p. 133). Thus in a single chapter of The American Voter the use of a candidate's party label is described both as a clear example of instrumental rationality and as an instance of ignorant conformity to social pressures. Later the authors introduce the notion of "levels of conceptualization," consigning other possible cues such as group benefits, the "nature of the times" under given parties, and "ideology-by-proxy" also to the realm of the sub-rational ([1960], p. 220).

Converse [1966] himself presents a sophisticated model of the use of information to update prior beliefs, but, given low levels of political information in the public, he also emphasizes the non-rational nature of voting. He makes the following analogy of political information with mass and momentum:

> ... the probability that any given voter will be sufficiently deflected in his partisan momentum to cross party lines ... varies directly as a function of the strength of short-term forces toward the opposing party and varies inversely as a function of the mass of stored information about politics (Converse [1966], p. 141).

The empirical result of this information-gathering process is that low-involvement voters are highly susceptible to change if any information does reach them, but vote strictly for party (if at all) in the majority of cases in which they remain uninformed. But Converse certainly does not take this as an indication of rational behavior by voters. He demonstrates (Converse [1964]) that most voters are completely uninformed on most issues, their responses to survey items being largely random rather than being based on informed preferences; a fortiori, voters are not competent to make coherent ideological connections between issues, connections which (Converse claims) are the stuff of elite-level politics. This view of the possibly incompetent voter is seconded by Stokes and Miller [1966], who show that voters have almost none of the information on congressional politics that would allow them to cast issue-based votes for congressmen.

Thus even among the voting behavior students of the Columbia and Michigan schools, who view the voter as being either inherently nonrational or too abjectly ignorant to cast a rational vote, the concept of essentially rational use of imperfect information makes a fairly distinct appearance. It was those scholars' failure to sufficiently combine that rationality with voting and with democratic politics that aroused Schattschneider's [1960] ire. However, it is Downs [1957] who provides the original, definitive counterpoint to their view.

2. THE RATIONALLY UNINFORMED VOTER

Using an explicitly rational-actor approach to model voting behavior, Downs ([1957], p. 209) defines the vote and the information-gathering process as part of a single problem for the voter: the "voting decision" consists of gathering and evaluating information, deciding whom to favor, and then either voting or abstaining. Downs explicitly assumes information to have a "cost" (p. 209) in terms of a diversion of resources, especially time. The amount of information to be gathered is based on a preliminary estimate of party differential and the "estimated cost of being wrong" (p. 241). Then,

> the costs and returns of all data must be weighed and information procured only if its expected return exceeds its cost. ... [I]nformation ... can be translated into a ... change in the utility income [the voter] expects if one of the parties is elected (p. 241).

Although Downs does not explicitly make use of Bayesian updating (it is unnecessary given his nonmathematical exposition), his concept of the updating process goes beyond those of the Columbia and Michigan scholars in two ways: he clearly realizes the separate roles of party differential, variance of the prior distribution, and accuracy of the information (Downs [1957], pp. 242-243); and he realizes the interdependency between rational voting choice and rational information gathering.

Downs ultimately fails, however, to present a complete, consistent model of the rational voter (see Barry [1970], pp. 99-118) because he assumes that the voter votes only in order to possibly influence the outcome of the election, using as a decision criterion the maximization of expected value of the voting-nonvoting choice minus the costs of information and voting. Because in a large society the probability of influencing the outcome is essentially zero, it will never be rational under these assumptions to gather information or to vote. Downs recognizes this impasse and grapples with it unsuccessfully ([1957], pp. 244-245). While this problem is easily surmounted by alternative concepts of rational voting (see Ferejohn and Fiorina [1974], as well as Chapter 2 of this study), its consideration leads Downs to two further interesting conclusions. First, he notes that voters who have interests aside from voting which require gathering political information may have more political power as a result ([1957], pp. 256-257), mentioning as an example the advantage of organized economic interests over consumers and taxpayers. Bartlett [1973] expands considerably on exactly this theme in his Downsian-style analysis of political power and society as a whole. Second, Downs suggests that political parties may offer "ideologies" as a cheap informational cue to voters ([1957], pp. 96-113). Although his version of ideology seems rather at odds with usual definitions, the use of it by Downs' artificial, perfectly-coordinated parties strongly resembles the concept of party identification as an informational cue which we noted in the Michigan literature. Here, however, its function is strictly rational and integrated into the whole process of rational voting.

Popkin et al. [1976] have elaborated upon Downs' model of the voter, especially upon the function of party and ideology as cues. They view each vote as "an investment in one or more collective goods under conditions of uncertainty with costly and imperfect information" (Popkin et al. [1976], p. 780). Notice that Downs' problem with the paradox of nonvoting would be expressed here as a possible free-rider problem, depending (just as for Downs) on the decision criterion of the voter. Again,

> An instrumental voter applies information ... which bears upon the expected returns to the voter of the election of particular candidates. The accumulation of information always involves the expenditure of resources by individuals.
>
> ... The investor-voter will use partisan and ideological labels as practical solutions to the problem of costly information (Popkin et al. [1976], p. 787).

This view of party "identification" as a rational voter's strategy for economizing on information is shared by Robertson [1976], who presents a careful critique of the group identification concept and of the contrary notion of some European scholars that party identification is wholly artificial (see Budge et al. [1976]). Popkin et al. [1976] suggest another cue as well: voters may use candidate characteristics such as features of their backgrounds and personalities to infer their desirability on policy performance (pp. 792-795), a possibility also suggested by Kelley ([1960], p. 12). Candidate and party orientations both appear as components of the electoral decision in Campbell et al. [1960], but their psychological "field-of-forces" model does not specify whether these are to be understood as cues for issue orientation, which is a separate component.

Thus Downs, Popkin, and others taking an "economic" approach do see the voter as having preferences over government policies. In spite of imperfect information, their voters strive to vote according to these preferences as far as is consistent with overall rational behavior. In the process, voters may economize by using party and other cues, and will certainly make their final choices with less than complete information. One important implication of such a theory is that even if voters are rational, we may observe such phenomena as party identification and uninformed voters; but as long as voters' cues are not on balance misleading, the tendency is to vote one's preferences, and electoral outcomes will in some degree reflect these preferences. Downs and his successors have described a scenario in which "ordinary people" might control a democratic government. However, they do not for the most part examine rigorously the logical properties of the voter's choice problem which they have proposed. This in fact turns out to be a nontrivial endeavor, not only for students of voting but for theorists of other rational choice problems as well. We turn next to a fruitful approach to this question as it appears in the literature of statistical decision theory, operations research, and economics.

3. INFORMATION AND PROBLEMS OF SEQUENTIAL CHOICE IN ECONOMICS

It is by now widely accepted among economic theorists that the behavior of economic agents and systems under uncertainty differs fundamentally from that under conditions of full information. As McCall [1971] points out, it is insufficient to replace marginal cost and marginal utility with "expected marginal cost" and "expected marginal utility," and hence simply apply traditional deterministic models, because there are phenomena observable under conditions of uncertainty which cannot be explained by such models. The most straightforward of these rest on the notion of agents' risk preference; for example, a risk-averse individual will choose an alternative of lower expected value if it has sufficiently lower risk, and people are willing to buy insurance for amounts greater than the expected loss from the events insured against. Also, under uncertainty, information takes on a positive value, a phenomenon which obviously does not appear in models assuming full information. Finally, the "irreversibility" of decisions and the need for flexibility to respond to unforseeable changes in the state of nature can have unexpected effects upon optimal behavior (see for example Arrow and Fisher [1974]).

The analytical tools for a large class of models of economics under imperfect information relevant to the present study were presented early on, by Wald [1950] and MacQueen and Miller [1960] among others. However, the seminal economic models are those of Stigler [1961, 1962] who suggests that rational behavior of consumers or job seekers must include rational search for information about prices and wage offers; Stigler models the economic agent as choosing, in advance, an optimal number of costly observations to take before stopping to choose one of them. McCall [1970] points out that this optimal sample size approach would be inferior, for the agent, to a process of sequential search in which the agent would decide after each observation whether to stop taking further observations. McCall's model of sequential search for wage offers begins a whole literature of optimal information-seeking behavior of buyers and job seekers. Rothschild's [1974] model of searching for the lowest price when the distribution of prices is unknown comes closest in method to that used in the present study. Finally, a seminal article of broader applicability than the sequential search models is Hirshleifer's [1971] exposition on the value of information in general economic choice problems.

Recent work in economics has turned more directly to the explanation of aggregate-level phenomena using individual behavior under uncertainty. First, there is the problem of modelling entire markets rather than just individual behavior. Rothschild [1973] and Hirshleifer [1973] point out that such models as those of Stigler [1962] and McCall [1970] are <u>less</u> than partial equilibrium analyses: the assumption that individual job seekers face a distribution of wage offers is not related to any theory of the motives of employers. In fact, a full market of the type described by Stigler or McCall would exhibit "unraveling" of the wage distribution in which the distribution would collapse, at equilibrium, to a single value. Since it is just such phenomena as multiple-valued equilibria that Stigler first sets out to explain, the original models require further improvement. Mortensen [1976], Wilde [1977], and Butters [1977] have all presented full, two-sided market models which accommodate such nondegernate equilibria.

A second aggregate-level question involves the idea of economic efficiency when information is incomplete. Arrow [1963] first suggested that underproduction of information alone could render a market outcome inefficient from the standpoint of social welfare. More abstract analyses by Spence [1974a, 1974b, 1976] examine Pareto efficiency and welfare maximization under several models of job search in which both information seeking and "signalling" take place.

Although the analysis of electoral politics under uncertainty is essentially at the level of Stigler [1962] and McCall [1970], such questions of aggregate effects and two-sided markets are relevant there also. One important goal should be to trace the effects of information-economizing behavior by voters on candidate strategies, which is conceptually already a concern of political science (see for example Kelley [1960]). Since both voters and politicians face uncertainty, a similar problem for political scientists is to understand the results in electoral outcomes and government policy choices. Here we again face the problem identified by Schattschneider [1960], to explain the workings of democracy under the realities of ignorance and uncertainty.

4. FORMAL MODELS OF VOTING BEHAVIOR UNDER UNCERTAINTY

More rigorous treatments of the problem of voting under costly and imperfect information than that of Downs [1957] have only just begun to appear. Shepsle [1972] first begins to model this process when he attempts to discuss the ambiguous position-taking of candidates. He notices that the risk preference of voters will determine whether they prefer a more ambiguous or a less ambiguous candidate. Tollison and Willett [1973] use a model consistent with Bayesian updating to draw the connection between more information and more precision in the voter's estimate of candidate positions, again noting that the voter's attitude toward risk helps determine whether more information will make the difference in expected utility between candidates larger or smaller. A different tack is taken by Fiorina [1977], who argues that past performance of the parties is a source of cheap and reliable information about future performance. His model of voting and the development of party identification uses both past and present information to determine the voter's orientation. We will discuss Fiorina's model in considerable detail in Chapter 3.

The first full-scale, decision-theoretic model of voting and information gathering is that of Zechman [1978]. Concerned with how rational voters might use past performance of the parties as information and the implications of this for the mobility of parties in an issue space, Zechman uses a formal Bayesian model to allow his voters to update past party preferences. The voter is assumed to have a prior distribution over each party's location in an issue space, and to use the party's current platforms as single observations to update the priors for the current election. The resulting posterior distribution becomes the prior at the next election. Such a Bayesian setup is alone sufficient to account for both the use of past performance and for possible spatial immobility of the parties, not immobility of actual position but rather of the positions perceived by the voters. As Zechman points out (p. 31, footnote), this model is highly compatible with Fiorina's [1977] model. Using simulation models, Zechman demonstrates some implications of such a model for party strategies, suggesting that the advertising of policy positions and advertisement of actual changes in position are distinct elements of the campaign process.

Zechman's concept of the development of party preferences is also consistent with several of the observed properties of party identification: for example, longer favorable experience with a party leads to stronger preferences for that party, and party preference is not as volatile as the preference between current party platforms. However, because he uses essentially the Downsian model of parties as teams, rather than distinguishing current candidates from parties, further interesting predictions are overlooked: for example, the possibility of voting against one's party identification without changing that identification. Keeping separate subjective distributions for parties and for current candidates would thus add further explanatory power, but the model is obviously the same in its essentials. Zechman's simplification does force him into one awkward theoretical device, grafting an ad hoc parameter which describes a voter's "receptivity to the possibility that Party i's future program has changed since the last election" (p. 28) onto the otherwise straightforward Bayesian model.

While it represents a definite step forward in modelling rational voting, Zechman's theoretical approach is limited in two important ways. First, he assumes for mathematical convenience that all priors and sampling distributions are normal with known precision. But the normal distribution with unknown mean is its own conjugate prior, and the Bayesian process in this case has an unusually simple form (see DeGroot [1970], p. 167). In particular, the updated mean is just a weighted average of the observation and the prior mean, the new precision is just the sum of the prior precision and the sampling precision (and hence always increases), and the posterior is still a normal distribution. There is no assurance that other distributions will behave in any such fashion under Bayesian updating, and there is as a result no reason to believe that Zechman's results are robust to relaxations of these assumptions. Second, Zechman's model is not a model of the process of rational information gathering. Only one kind of observation is available, exactly one time, and it is costless. Thus much of the variation in voter behavior which may appear when different voters face different sampling distributions at different, nonzero costs cannot appear in the model. In the context of the present study, Zechman's model describes the updating part of the process, but not the rational use of costly information. Party label is used by voters as a cue exactly because it is the only cue available; Zechman tells us where a rational voter's party identification might come from, but we still must learn why and how it is related to voting.

In a research note, Coughlin [1976] sets out to address "the definition and analysis of an election process with incomplete information and noisy signals from political candidates" (p. 113), and indeed he defines terms which cover all the necessary ingredients of a full-rationality, two-sided electoral model of candidate signalling and voter sequential search. However, these definitions represent the full extent of the model. His "analysis" shows only that as a direct result of the definitions such an election represents a two-person, zero-sum game between the two candidates. He claims that existence of optimal ambiguous strategies is assured by the Minimax Theorem (p. 115), but even this is false in general unless one adds the requirement that the issue space has only finitely many elements.

To date, the only formal analysis of the combined voting and information problem is that of Powell [1974]. Her analysis applies to any voter objective function which is a linear function of the difference in actual utility between the two candidates. Before either choosing a candidate or abstaining, the voter is permitted to sample noisy observations of the true utility of each candidate by paying a search cost. Starting from a uniform prior distribution of the utility difference, the voter incorporates his observations using Bayes' rule and makes the final choice which minimizes the expected regret of the final choice. This objective is claimed by Powell to be equivalent to maximizing expected utility, but she does not prove this fact; we supply a proof in the Appendix to this chapter. Although Powell's model deals explicitly with voter behavior in a single election, it is conceptually a completion of Zechman's [1978] model of voter choice: besides engaging in Bayesian updating, the voter must gather information to whatever extent is appropriate, given his own preferences and the kinds of information available.

Having set up the complete model, Powell performs a comparative statics analysis to examine the effects, on optimal sampling behavior and optimal choice of candidates (including abstention), of changes in the variance of the sampling distribution, the cost function for sampling, and the true utility difference between the candidates. This analysis is accomplished mainly through differentiation of the posterior expected regret function with respect to costs, variance, and sample size. Also, it is observed that the variance of the candidate utilities and the amount of information collected are irrelevant in the determining of the final optimal choice (Powell [1974, p. 14); but this is of course an artifact of the model's use of direct observation of utility values, bypassing the problem of the voter's risk preference. A fuller discussion of this issue is pursued in Chapter 3.

In identifying the optimal decision for comparative statics analysis, Powell uses an optimal sample size approach like that of Stigler [1962]. She contends (Powell [1974], pp. 21-22) that this is more realistic than a sequential analysis, since the voter is likely to be unable (even in principle) to construct an optimal sequential-sampling plan, while the calculations necessary for finding the optimal number of observations before any sampling takes place are at least amenable to closed-form solution. Even so, the voter must supply estimates of true parameter values in order to perform this calculation before any updated estimates are available. At any rate, the closed-form character of the optimal sample size does lend itself to conventional comparative statics analysis; at the same time, it presents a serious drawback. The analysis begins with setting the derivative of expected regret, with respect to sample size, equal to zero, and this identity underlies the whole derivation.

But sample size is of course an integer value in reality, so that at the optimum integer value, the aforementioned derivative may not be zero. It is not immediately clear how this discrepancy carries through the analysis, and thus it is difficult to claim that the results are close to the results given the actual integer solutions. Stigler [1962] is of course open to exactly these sorts of criticisms as well.

Another problem of differentiation creates a much more serious difficulty in Powell's analysis. Let h be the posterior density of the utility difference $U_1 - U_2$, f the sampling density, and g the prior. For a given sample of utility values, let $\overline{U}_1 - \overline{U}_2$ be the difference in the means of the observations. Powell chooses to then write Bayes' rule as

$$h(U_1 - U_2 | \overline{U}_1 - \overline{U}_2) = \frac{g(U_1-U_2)F(\overline{U}_1-\overline{U}_2 | U_1-U_2)}{\int_{-\infty}^{\infty} g(t)f(\overline{U}_1-\overline{U}_2 | t)dt}$$

(Powell [1974], p. 11). Of course for a given sample size, this is valid; but the sample size and thus the variance of $\overline{U}_1 - \overline{U}_2$ become parameters of f, and hence of h. That is, the way $\overline{U}_1 - \overline{U}_2$ enters f depends on the sample size and the sample variance. Since the posterior expected regret includes the posterior density h, any differentiation with respect to sample size or variance of the sample mean must take this parameter as an arguemnt of h; but this is not done in Powell's analysis (see Powell [1974], pp. 17, 19-20 for the relevant calculations). It appears to be the form of Bayes' rule given above which allows closed form solution of the voter's optimal sample size problem; taking into account the effects of sample size on the updating formula may render this analysis intractable as well. As given, the derivatives calculated, upon which the comparative statics results depend, are incorrect, and cannot be corrected by any simple manipulations of the formulas given.

Thus the study of rational voting with costly and imperfect information has barely started. Few rigorous models have even been suggested, fewer results have been obtained, and the possibility of a full model of candidates and voters remains unexamined. In the present study we will contribute to the question of voting behavior, but a less direct approach to aggregate effects remains a necessity.

5. THE PRESENT STUDY: OVERVIEW

In order to address Schattschneider's [1960] problem of determining how ordinary citizens can influence democratic government, this study approaches the voter's problem as one of rational, sequential search for information with which to choose between risky alternatives. First, however, there is the matter of the voter's preferences. No matter how the voter informs himself, he cannot help assure responsible government unless his vote is ultimately related to the issues and outcomes of government policy. Models assuming expected value maximization of a voter's direct effect on policy have run aground on this problem, either being inconsistent on the question of nonvoting (Downs [1957]) or attributing voting to a "citizen duty" term not based on political issues (Riker and Ordeshook [1968]). Either way, it does not pay the rational voter to gather costly information about the candidates' issue positions. In this study we accept the contention of empirical researchers such as Key [1966], Pomper [1972], and Boyd [1972] that votes are related to policy preferences; we employ the most general model possible which is consistent with their observations and with rational choice, by assuming that the voter's utility depends on the actual positions of the candidates and upon his own vote. This approach is consistent with any issue-related voting, be it expressive or instrumental, investment or consumption (approaches distinguished by Fiorina [1976]).

In Chapter 2 we apply this structure of voter preferences to a standard issue space analysis. As in Zechman [1978], the voter observes issue positions, this time of candidates in a single election, updating his priors about the candidates' true positions. In contrast with Powell [1974], we model the information-gathering process as one of sequential search: although it is true that the optimal sample-size approach offers easier calculation, it is unreasonable to assume that voters make <u>any</u> such calculation. But the optimal sequential search process asks, before each observation, the very question that a rational human actor might be expected to ask: is it worth the trouble to take this observation? The formal analysis of sequential search then simply analyzes what will be the nature of such a process, if the actor always answers this question in a manner consistent with his preferences and expectations.

Formally, the problem is one of dynamic programming. To answer whether indeed there can be a consistent, rational course of behavior for voters in such a model, we demonstrate that this dynamic programming problem has a well-defined optimal solution. This is done by proving (Theorems 1 and 2) that the corresponding functional equation has a unique solution; hence an optimal sampling plan exists. We proceed to examine some properties of optimal sampling behavior: its response to changes in sampling costs (Theorems 3 and 4) and to changes in the risk environment (Theorem 5). Some connections between the outcome of the sampling process and the voter's choice of candidates are then examined (Theorem 6 and the ensuing counterexamples). Finally, we explore possible extensions of the model to include multicandidate elections, nonvoting due to indifference and alienation, and multiple sampling distributions.

In Chapter 3 we proceed to a variant of the model, in which voters observe the possible values of the outcomes of candidates' policies but not the policy positions themselves. This viewpoint has ample precedent in the study of electoral behavior, and its connections with a formal, non-rational model of Fiorina [1977] are examined in some detail, along with other models along these lines. Whereas Zechman's [1978] model could explain the updating of party identification in Fiorina's [1977] model as a Bayesian process, the present model allows us to understand Fiorina's model much more fully as a process of rational behavior under imperfect information. As is pointed out, empirical tests of the Fiorina model may consequently be understood as tests of the voter search model of the present study.

In this study, then, we define a process through which ordinary citizens, faced with the uncertainty and ignorance which typify life in a complex world, may nevertheless transmit their wishes to a democratic government. The significance of these results, as theory, rests on the fact that the model covers all aspects of the individual's voting behavior. But what about the other side of the system, the democratic government itself? It is necessary also for political scientists to say whether government can or will respond to such electoral signals in such a way that voters' wishes are reflected in government policy. Although this is a difficult theoretical problem, we attempt in this study to examine indirectly some of the implications of the voting model for electoral outcomes.

6. ELECTION OUTCOMES UNDER UNCERTAINTY: SURVEY AND PREVIEW

In the Appendix to Chapter 4 we demonstrate how the voter search model of Chapter 2 can be used directly to improve certain spatial models of candidate competition, typified by Hinich et al. [1972], in which various convexity and concavity assumptions are employed to guarantee the existence of equilibrium candidate strategies. Apart from this application, the state of the art restricts us to indirect use of the notions of voter uncertainty to study election outcomes, particularly through the effect which candidate uncertainty (partly a result of voter uncertainty) may have.

Downs [1957] points out the possible implications for candidate strategies of the imperfect information of voters. He suggests that political parties will offer ideological labels to serve as informational cues for voters, labels which may place some constraints on the parties' actual platforms. In addition, Downs notes the advantage in political influence that will accrue to organized interests who can afford to gather large amounts of information; the parties will be more sensitive to the needs of these more informed interests. More formal analyses of the situation are necessarily less ambitious. Shepsle [1972] notes that candidates may present ambiguous positions to the voters; assuming rational voting in the face of this uncertainty, Shepsle notes that according to the distribution of risk preference in the electorate, more ambiguous or less ambiguous strategies may be advantageous to the candidates. Page [1976] presents a thoughtful critique of Shepsle's work and develops further the theory of political ambiguity in a less rigorous fashion. Of course, candidates themselves may be uncertain about election outcomes; Wittman [1975] and Denzau and Kats [1977] have examined the optimal strategies for candidates whose objective functions are probabilistic. We discuss these two works more fully in Chapter 4, and discuss the nature and sources of candidate uncertainty.

The main subject of our analysis of electoral outcomes under uncertainty is the connection between informational assumptions and goal assumptions in modelling candidate behavior. When candidates seek only to gain office, the kinds of uncertainty discussed have no great effect on predicted outcomes. Suppose, however, that the candidates are interested in the policies they will implement aside from their function in attracting votes; then the presence of uncertainty about electoral outcomes leads to a prediction that candidate platforms will not converge in equilibrium as they do in conventional spatial models, either with candidates whose only goal is winning or with policy-oriented candidates who are _not_ uncertain about outcomes (the latter are modelled in Wittman [1973, 1977] as well as in Chapter 4).

Thus when candidates are responsive to the vote, the extent of uncertainty may determine how they respond to the vote. It must remain for later studies to examine more directly the connection between the strategies of candidates and officeholders and the information-economizing behavior of voters. Both theoretical issues, such as how rational candidates might take advantage of voters' information gathering, and empirical questions, such as how different kinds of information about policies impinge on the voters' awareness, are of obvious significance in understanding the democratic process. Certainly the voter and candidate models of the present study are a step in that direction.

Appendix

EXPECTED UTILITY MAXIMIZATION AND EXPECTED REGRET MINIMIZATION

Consider the following general decision problem under risk: the true state of the world is known only as a random variable X with distribution function F. W is a set of possible acts, one of which is to be chosen by a decisionmaker (W may be infinite). A utility function u is defined over states of the world and actions chosen: utility is $u(x,w)$ if $X = x$ is realized and $w \in W$ is chosen. Let the <u>regret function</u> be defined over the same pairs by

$$R(x,w_0) = \sup_{w \in W} u(x,w) - u(x,w_0).$$

The expected utility of an act is given by

$$Eu(w) = \int u(x,w)dF(x)$$

and the expected regret by

$$ER(w) = \int R(x,w)dF(x),$$

which is assumed to exist.

Theorem: Let a,b be two acts in W. Then $ER(a) \geq ER(b)$ if and only if $Eu(a) \leq Eu(b)$.

Proof: Suppose $Eu(a) \geq Eu(b)$. Let $M(x) = \sup_{w \in W} u(x,w)$. Let

$$A = \{x \mid u(x,a) = M(x) > u(x,b)\}$$

$$B = \{x \mid M(x) > u(x,a) > u(x,b)\}$$

$$C = \{x \mid M(x) > u(x,b) > u(x,a)\}$$

$$D = \{x \mid M(x) = u(x,b) > u(x,a)\}$$

$$S = \{x \mid u(x,a) = M(x) = u(x,b)\}$$

$$T = \{x \mid M(x) > u(x,a) = u(x,b)\}.$$

Then

$$\int_{A \cup B \cup C \cup D \cup S \cup T} u(x,a)\,dF = Eu(a)$$

$$\geq Eu(b) = \int_{A \cup B \cup C \cup D \cup S \cup T} u(x,b)\,dF$$

so

$$\int_{A \cup B} [u(x,a) - u(x,b)]\,dF + \int_{S \cup T} u(x,a)\,dF$$

$$\geq \int_{C \cup D} [u(x,b) - u(x,a)]\,dF + \int_{S \cup T} u(x,b)\,dF,$$

or, since $u(x,a) = u(x,b)$ on $S \cup T$,

(*) $\int_{A \cup B}[u(x,a) - u(x,b)]dF \geq \int_{C \cup D}[u(x,b) - u(x,a)]dF.$

The following relations hold for R:

(i) on A, $R(x,a) = 0$ and $R(x,b) = u(x,a) - u(x,b)$;

(ii) on B, $R(x,a) = M(x) - u(x,a)$ and

$$\begin{aligned} R(x,b) &= M(x) - u(x,b) \\ &= M(x) - u(x,a) + u(x,a) - u(x,b) \\ &= R(x,a) + [u(x,a) - u(x,b)]; \end{aligned}$$

(iii) on C, $R(x,b) = M(x) - u(x,b)$ and

$$\begin{aligned} R(x,a) &= M(x) - u(x,a) \\ &= R(x,b) - [u(x,b) - u(x,a)]; \end{aligned}$$

and

(iv) on D, $R(x,b) = 0$ and $R(x,a) = u(x,b) - u(x,a)$.

Hence substituting in (*) and rearranging,

$$\begin{aligned} &\int_A R(x,b)dF + \int_B [M(x) - R(x,a) - M(x) + R(x,b)]dF \\ &\geq \int_D R(x,a)dF + \int_C [M(x) - R(x,b) - M(x) + R(x,a)]dF. \end{aligned}$$

Cancelling and collecting terms,

$$-\int_{B \cup C \cup D} R(x,a)dF \geq -\int_{A \cup B \cup C} R(x,b)dF.$$

Since $\int_S R(x,a)dF = \int_S R(x,b)dF = 0$ and $\int_T R(x,a)dF = \int_T R(x,b)dF$ we now have

$$-ER(a) \geq -ER(b)$$

$$ER(a) \leq ER(b). \qquad \square$$

The significance of the Theorem is this: whether the decision-maker uses expected utility maximization or what we may call expected regret minimization, his ranking of alternative actions in W will be the same. Thus we have the following important result:

<u>Corollary</u>: For a general decision problem under risk as characterized here, expected utility maximization and expected regret minimization lead to the same choice of acts.

However, expected utility maximization and its equivalent, expected <u>loss</u> minimization, are not on the surface equivalent to expected regret minimization, since the regret function is not the same as the loss function (compare Powell [1974]).

REFERENCES

Almond, Gabriel A. and Sidney Verba. The Civic Culture. Princeton, N.J.: Princeton University Press, 1963.

Arrow, Kenneth J. "Uncertainty and the Welfare Economics of Medical Care." American Economic Review 53 (1963): 941-973.

Arrow, Kenneth J. and Anthony C. Fisher. "Environmental Preservation, Uncertainty, and Irreversibility." Quarterly Journal of Economics 88 (1974):312-319.

Barry, Brian M. Sociologists, Economists, and Democracy. London: Macmillan, 1970.

Bartlett, Randall. Economic Foundations of Political Power. New York: Free Press, 1973.

Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee. Voting. Chicago: University of Chicago Press, 1954.

Boyd, Richard W. "Popular Control of Public Policy: A Normal Vote Analysis of the 1968 Election." American Political Science Review 66 (1972):429-449.

Budge, Ian, Ivor Crewe, and Dennis Farlie, eds. Party Identification and Beyond. London: John Wiley and Sons, 1976.

Butters, Gerard. "Equilibrium Distributions of Sales and Advertising Prices." Review of Economic Studies 44 (1977): 465-491.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. The American Voter. New York: John Wiley and Sons, 1960.

Converse, Philip E. "The Nature of Belief Systems in Mass Publics." In Ideology and Discontent, edited by David E. Apter. New York: Free Press, 1964.

________________. "Information Flow and the Stability of Partisan Attitudes." In Elections and the Political Order, edited by Angus Campbell, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. New York: John Wiley and Sons, 1966.

Coughlin, Peter. "Incomplete Information, Noisy Signals, and Uncertainty-Averse Voting in Political Elections: A Note." Public Choice 28 (Winter 1976):113-116.

DeGroot, Morris H. Optimal Statistical Decisions. New York: McGraw-Hill, 1970.

Denzau, Arthur T. and Amoz Kats. "Expected Plurality Voting Equilibrium and Social Choice Functions." Review of Economic Studies 44 (1977):227-233.

Downs, Anthony. An Economic Theory of Democracy. New York: Harper and Brothers, 1957.

Ferejohn, John A. and Morris P. Fiorina. "The Paradox of Not Voting: A Decision Theoretic Analysis." American Political Science Review 68 (1974):525-536.

Fiorina, Morris P. "The Voting Decision: Instrumental and Expressive Aspects." Journal of Politics 38 (1976): 390-413.

______________. "An Outline for a Model of Party Choice." American Journal of Political Science 21 (1977):601-626.

Hinich, Melvin J., John O. Ledyard, and Peter C. Ordeshook. "Nonvoting and the Existence of Equilibrium Under Majority Rule." Journal of Economic Theory 4 (1972): 144-153.

Hirshleifer, Jack. "The Private and Social Value of Information and the Reward to Inventive Activity." American Economic Review 61 (1971):561-574.

______________. "Where Are We in the Theory of Information?" American Economic Review 63 (1973):31-39.

Kelley, Stanley Jr. Political Campaigning. Washington: The Brookings Institution, 1960.

Key, V.O. Jr. The Responsible Electorate. Cambridge, Mass.: Belknap Press of Harvard University Press, 1966.

Lazarsfeld, Paul F., Bernard R. Berelson, and Hazel Gaudet. The People's Choice. 2nd rev. ed. New York: Columbia University Press, 1948.

Lipset, Seymour Martin. Political Man. Garden City, N.Y.: Doubleday, 1960.

MacQueen, J. and R.G. Miller, Jr. "Optimal Persistence Policies." Operations Research 8 (1960):362-380.

Margolis, Michael. "From Confusion to Confusion: Issues and the American Voter (1956-1972)." American Political Science Review 71 (1977):31-43.

McCall, John J. "Economics of Information and Job Search." Quarterly Journal of Economics 84 (1970):113-126.

____________. "Probabilistic Microeconomics." The Bell Journal of Economics and Management Science 2 (1971):403-433.

Mortensen, Dale T. "Job Matching under Imperfect Information." In Evaluating the Labor Market Effects of Social Programs, edited by O. Ashenfelter. Princeton, N.J.: Princeton University Press, 1976.

Page, Benjamin I. "The Theory of Political Ambiguity." American Political Science Review 70 (1976):742-752.

Pomper, Gerald M. "From Confusion to Clarity: Issues and American Voters, 1956-1968." American Political Science Review 66 (1972):415-428.

Popkin, Samuel, John W. Gorman, Charles Phillips, and Jeffrey A. Smith. "Comment: What Have You Done for Me Lately? Toward an Investment Theory of Voting." American Political Science Review 70 (1976):779-805.

Powell, Lynda Watts. "A Bayesian Model of Individual Voting Decisions in Two-Candidate Elections." Mimeographed. Albany, N.Y.: Graduate School of Public Affairs, State University of New York at Albany, 1974. (Forthcoming in Journal of Mathematical Sociology 6 (1978).)

RePass, David E. "Issue Salience and Party Choice." American Political Science Review 65 (1971):389-400.

Riker, William H. and Peter C. Ordeshook. "A Theory of the Calculus of Voting." American Political Science Review 62 (1968):25-42.

Robertson, David. "Surrogates for Party Identification in the Rational Choice Framework." In Party Identification and Beyond, edited by Ian Budge, Ivor Crewe, and Dennis Farlie. London: John Wiley and Sons, 1976.

Rothschild, Michael. "Models of Market Organization with Imperfect Information: A Survey." Journal of Political Economy 81 (1973):1283-1308.

___________. "Searching for the Lowest Price When the Distribution of Prices Is Unknown." Journal of Political Economy 82 (1974):689-711.

Schattschneider, E.E. The Semisovereign People. New York: Holt, Rinehart, and Winston, 1960.

Shepsle, Kenneth A. "The Strategy of Ambiguity: Uncertainty and Electoral Competition." American Political Science Review 66 (1972):555-568.

Spence, Michael. "Competitive and Optimal Responses to Signals: An Analysis of Efficiency and Distribution." Journal of Economic Theory 7 (1974a):296-332.

__________. Market Signaling. Cambridge, Mass.: Harvard University Press, 1974b.

__________. "Competition in Salaries, Credentials, and Signaling Prerequisites for Jobs." Quarterly Journal of Economics 90 (1976):51-74.

Stigler, George J. "The Economics of Information." Journal of Political Economy 69 (1961):213-225.

__________. "Information in the Labor Market." Journal of Political Economy 70 (1962):94-105.

Stokes, Donald E. and Warren E. Miller. "Party Government and the Saliency of Congress." In Elections and the Political Order, edited by Angus Campbell, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. New York: John Wiley and Sons, 1966.

Tollison, R.D. and T.C. Willett. "Some Simple Economics of Voting and Not Voting." Public Choice 16 (Fall 1973):59-71.

Wald, Abraham. Statistical Decision Functions. New York: John Wiley and Sons, 1950.

Wilde, Louis L. "Labor Market Equilibrium under Nonsequential Search." Journal of Economic Theory 16 (1977):373-393.

Wittman, Donald A. "Parties as Utility Maximizers." American Political Science Review 67 (1973):490-498.

______________. "Candidate Equilibrium." Mimeographed. Santa Cruz, Calif.: Department of Economics, University of California at Santa Cruz, 1975.

______________. "Candidates with Policy Preferences: A Dynamic Model." Journal of Economic Theory 14 (1977):180-189.

Zechman, Martin J. "Dynamic Models of Voting Behavior and Spatial Models of Party Competition." Institute for Research in Social Science Working Papers in Methodology, Number 10. Chapel Hill, N.C.: University of North Carolina, 1978.

Chapter 2

THE VOTER'S SEQUENTIAL SEARCH PROBLEM

The general choice problem of the voter is to choose the candidate whose issue positions are ranked highest in the voter's preference ordering. In this chapter we consider how this choice is complicated when the voter cannot be certain what a candidate's position really is. When information about those positions is costly and is itself subject to error, the voter is faced with a choice of information-gathering strategies as well as the risky choice between candidates.

Of course, there are several ways to operationalize, for purposes of modelling, the voter's objectives. The conceptual model underlying the present analysis is as follows: the voter is assumed to have preferences over consumption bundles which include the position y of the candidate he votes for, leisure L, and other private goods z. These preferences can be represented by a utility function $u^* = u^*(y,z,L)$ which is to be maximized. Notice that the utility of voting is neither directly instrumental (the voter need not consider his own influence on government policy) nor is it purely a matter of "citizen duty" (as in Riker and Ordeshook [1968]). Instead the voting act has a <u>consumption</u> value which depends upon the actual position of the candidate for whom the vote is cast.

On the face of it, this concept of voting is a reasonable one because it is consistent with (1) the observed tendency of many citizens to vote despite a lack of instrumental effectiveness (Riker and Ordeshook [1968]; Brody and Page [1973]) and perhaps independently of their expectations about the closeness of the election (Fiorina [1976]; Weisberg [1977]); and with (2) the observed dependence of vote upon expressed candidate preference and issue preference (Fiorina [1976]; Pomper [1972]; Boyd [1972]; RePass [1971]; Goldberg [1966]; Jackson [1975]; Shapiro [1969]). By giving up some of L and perhaps z, the voter can acquire information about the true values of y which are available. For the present we will focus on the choice of candidates and postpone the consideration of nonvoting.

Because the private goods likely to be given up for political information are not likely to be of such types or in such amounts that they will affect the voter's preferences over candidate positions, we will in effect assume that

$$u^*(y,z,L) = u(y) + u^{**}(z,L).$$

This formulation is useful for two reasons. First, it allows us to make use of the standard spatial modelling setup in which each voter has a preference function over policy positions alone, here expressed by u(y); and it allows us to express the cost of obtaining and evaluating information about y as a well-defined opportunity cost of foregoing other uses of leisure and private goods. This cost is now merely the reduction in $u^{**}(z,L)$ needed to trade z and L for information.

1. PRELIMINARIES

Let S be the set of all possible issue positions; assumptions such as boundedness or convexity of S will be necessary for certain purposes and will be made later. Let u be the individual's utility function on S. Candidates 1 and 2 are assumed to occupy positions y_1 and y_2 in S, but these are unknown to the voter. Instead, the voter has subjective probability density functions g_1 for candidate 1's position and g_2 for candidate 2's position. To the voter, then, the true candidate positions are independent random variables Y_1 and Y_2. We will write g for (g_1,g_2), a density function on the Cartesian product S x S. The voter is assumed to be an expected utility maximizer; if g is his prior, he calculates expectations $\underset{g}{E}u(Y_1)$ and $\underset{g}{E}u(Y_2)$. (Notice that we can write either

$$\underset{g}{E}u(Y_i) = \int_S u(y_i)g_i(y_i)dy_i$$

or

$$\underset{g}{E}u(Y_i) = \int_{SxS} u(y_i)g(y)dy$$

where $y = (y_1,y_2)$, since in the latter case we are just integrating g to get its marginal g_i. These expressions will be used interchangeably as is convenient.)

The voter's information gathering will be assumed to take the following form: at any time, the voter may take an observation of a random variable $X = (X_1, X_2) \in S \times S$ by incurring an opportunity cost c. These observations of the two candidates are assumed to be independent and to have density functions f_1 and f_2. (We will write f for the joint density of X_1 and X_2 except when the special case $f = f_1 = f_2$ is explicity assumed.) The simplifying assumptions, that observations come in pairs, from a single joint sampling distribution, at a constant cost, are made so that we can directly address the question of how much information the voter gathers without the complication of choosing between distributions from which to sample. We will discuss the relaxation of these assumptions in Section 8 of this chapter. The most important restriction implied here is that the sampling distribution or distributions are regarded by the voter as fixed. The significance of this assumption will be discussed in Chapter 5.

Finally, we place no restriction on the length of time required for an observation, or on the number of observations which may in principle be taken. This does not seem to be a serious modification, since voters' observation of candidates is probably limited more by information cost than by length of the campaign. Yet in considering the case where search cost is allowed to vary, we still need to think of the campaign as having finite duration; otherwise we would imagine the voter acquiring arbitrarily many observations by waiting for arbitrarily cheap ones.

The voter's observations are used to update his priors according to Bayes' Rule. If g is the prior density and an observation x is received, the updated prior will be written as $\psi(x,g)$. The values of $\psi(x,g)$ are then defined by

$$g(y|x) = \frac{g(y)f(x|y)}{\int g(t)f(x|t)dt}$$

where the integral is over S x S.

2. THE SEQUENTIAL DECISION PROBLEM

As formulated thus far, the task facing the voter is a <u>sequential decision problem</u> in the sense of DeGroot [1970]. The voter must choose a candidate so as to maximize expected utility given the voter's posterior probabilities. Whenever observation ceases and the choice is made, the voter's best alternative is to choose $i \in \{1,2\}$ to maximize $\underset{g}{E}u(Y_i)$ where g is the posterior density. At any stage, the voter may (1) stop observing and choose the best alternative, or (2) take another observation at a cost c. Let

$$v_0(g) = \max \{\underset{g}{E}u(Y_1), \underset{g}{E}u(Y_2)\}$$

represent the expected utility of stopping and choosing immediately when the subjective density function is g. To deal with the sequential decision problem, the voter must compare $v_0(g)$ with the expected utility of paying c, taking an additional observation, and proceeding in some fashion from there.

Following DeGroot [1970] we will denote a sampling plan by δ. This is a rule associating with every subjective probability density function g a decision $\delta(g)$, either "stop" (and choose the alternative with highest expected value) or "continue" (pay c and receive another observation $x = (x_1, x_2)$). Given a prior g and a sampling plan δ, we can define a random variable $K = K(g,\delta)$ whose value is the number of observations which will actually be taken. Its realization depends upon the actual observations $(x^1, x^2, \ldots)$ of pairs of candidate observations. Let us write $\xi = (x^1, x^2, \ldots) = [(x_1^1, x_2^1), (x_1^2, x_2^2) \ldots]$ for the actual value of the sequence of observations of X, which may in principle be arbitrarily long. Since the observations $X^1, X^2, \ldots$ are independent random variables each having distribution function F, the consistency theorem of Kolmogorov (see Feller [1971], page 123) guarantees that the random variable $(X^1, X^2, \ldots)$ has a distribution function Φ whose marginals are just F. Thus we can define the expected utility of the decision problem when g is the prior and δ the sampling plan by

$$v(g,\delta) = E[v_0\{\psi[(X^1, X^2, \ldots, X^K), g]\} + Kc]$$

where the expectation is taken with respect to Φ and is assumed to exist.

DeGroot asserts ([1970], p. 300) that for the search problem as stated here the following functional equation holds:

(1) $$v^*(g) = \max\{v_0(g), Ev^*(\psi(X,g)) - c\}.$$

In the remainder of this section we shall verify this functional equation form for the voter's sequential search problem, showing that a solution exists and that the solution is unique.

Let Γ be the set of all subjective probability density functions on S x S. For simplicity we will think of these as proper density functions with respect to Lebesgue measure, i.e., the corresponding probability measures are absolutely continuous with respect to Lebesgue measure. Most of the results can be generalized to include discrete probability functions. In the context of the voter's dynamic programming problem for sequential search, Γ is the set of <u>states</u>.

Let B be the σ-algebra over S x S over which the probability measures corresponding to f and g are defined. Let A_g be a σ-algebra on Γ such that $T \in B$ implies that $\{\hat{g} \in \Gamma \mid \hat{g} = \psi(x,g)$ for some $x \in T\} \in A_g$, and let A be the smallest σ-algebra containing $\bigcup_{g\in\Gamma} A_g$. We shall make the following measurability assumptions:

(i) $\Pr\{\psi(X,g) \in T\}$ is defined for all $T \in A$, for each g;

(ii) $v_0\colon \Gamma \to \mathbb{R}$ is a measurable function.

The following theorem appears in MacQueen and Miller [1960] as Theorem 1; we reproduce it here using the present notation:

<u>Theorem 1</u>: If v_0 is bounded on Γ then the equation

$$V(g) = \max\{v_0(g), EV(\psi(X,g)) - c\}$$

has at least one solution $V: \Gamma \to \mathbb{R}$.

<u>Proof</u>: Consider the corresponding finite-horizon dynamic programming problems:

$$\begin{aligned} V^0(g) &= v_0(g) \\ V^1(g) &= \max\{v_0(g), EV^0(\psi(X,g)) - c\} \\ V^2(g) &= \max\{v_0(g), EV^1(\psi(X,g)) - c\} \\ &\vdots \\ V^N(g) &= \max\{v_0(g), EV^{N-1}(\psi(X,g)) - c\}. \end{aligned}$$

It is easy to show by induction that $V^N(g) \geq V^{N-1}(g)$. Since clearly $V^N(g) \leq \sup_{h\in\Gamma} v_0(h)$ for all N and g, we know that $\lim_{N\to\infty} V^N(g) = V^*(g)$ exists and is measurable. Passing to the limit in the finite-horizon problem gives

$$\begin{aligned} V^*(g) &= \max\{v_0(g), \lim_{N\to\infty} EV^{N-1}(\psi(X,g)) - c\} \\ &= \max\{v_0(g), EV^*(\psi(X,g)) - c\}, \end{aligned}$$

the latter using the Lebesgue convergence theorem □

MacQueen and Miller [1960] proceed to give three proofs of the uniqueness of the solution to the functional equation. However, these depend on strong assumptions about the form of the solution, and on compactness of the state space. In Theorem 2 we show that merely assuming v_0 bounded and $c > 0$ is sufficient for uniqueness. This result is applicable to general sequential search/stopping rule problems.

Theorem 2: Suppose v_0 is bounded and $c > 0$. Then the functional equation (1) has at most one solution.

Proof: For any $R:\Gamma \to \mathbb{R}$ satisfying the functional equation, we make the following definitions:

(i) Let $\gamma_R:\Gamma \to \mathbb{R}$ be defined by $\gamma_R(g) = ER[\psi(X,g)] - R(g)$. Notice that $\gamma_R(g) \leq c$ for all g.

(ii) Let δ_R be a sampling plan $\delta_R: \Gamma \to \{\text{stop, continue}\}$ such that

$$ER[\psi(X,g)] - c > v_0(g) \Longrightarrow \delta_R(g) = \text{continue}$$

and

$$ER[\psi(X,g)] - c < v_0(g) \Longrightarrow \delta_R(g) = \text{stop}.$$

We define the expected payoff of a sampling plan δ at state g as follows: let K_δ be a random variable giving the number of times δ = continue, starting at g and updating before reaching a $\bar{g}$ where $\delta(\bar{g})$ = stop. The realization of K_δ is determined entirely by the realization of $(X^1, X^2, \ldots)$, the sequence of observations. Thus

$$v(g,\delta) = E\{v_0[\psi((X^1,\ldots,X^{K_\delta}), g)] - cK_\delta\}.$$

So by definition, for any R which satisfies the functional equation,

$$R(g) = v(g,\delta_R).$$

It can be seen that the boundedness of v_0 implies that any R satisfying the functional equation must be bounded:

$$v_0(g) > b \quad \forall g \implies \max\{v_0(g), ER[\psi(X,g)] - c\} > b$$

and

$$v_0(g) < B \quad \forall g \implies E\{v_0[\psi(X^1,\ldots,X^{K_\delta}),g] - cK_\delta\} < B - cEK_\delta$$

so $R(g) < B$ as well. Finally notice that if $\delta(g)$ = continue for all g (or with probability 1 for future updates) then $v(g,\delta) = -\infty$ due to the amassing of search costs; hence for such δ, $v(g,\delta)$ cannot satisfy the functional equation. Suppose V(g) and U(g) both satisfy the functional equation (1). Let $\Gamma_1 = \{g | V(g) = U(g)\}$, $\Gamma_2 = \{g | V(g) > U(g)\}$, $\Gamma_3 = \{g | V(g) < U(g)\}$. On Γ_3, it must be that δ_U = continue, because if $v_0(g) = U(g)$ then V(g) cannot be less than U(g) (since obviously $V(g) \geq v_0(g)$). (Notice that this implies that neither $\Gamma_1 \cup \Gamma_2$ nor $\Gamma_1 \cup \Gamma_3$ can be empty.) Thus $U(g) = EU[\psi(X,g)] - c$, i.e., $\gamma_U(g) = c$, hence $\gamma_V(g) \leq \gamma_U(g)$ for all $g \in \Gamma_3$.

The strategy of the remainder of this proof is to show existence of a $g^0 \in \Gamma_3$ for which $\gamma_V(g^0) > \gamma_U(g^0)$, a contradiction unless $\Gamma_3 = \emptyset$. Let $M = \inf_{g \in \Gamma}[V(g) - U(g)] = \inf_{g \in \Gamma_3}[V(g) - U(g)]$.

Notice $-\infty < M < 0$. (If Γ_3 is empty the proof is done.) We will make use of the following Lemma, to be proved later.

<u>Lemma</u>: There exists an $\varepsilon > 0$ and a λ, $0 < \lambda \leq 1$, such that for any $\mu > 0$ there is some state $g^0 \in \Gamma_3$ for which

(i) $V(g^0) - U(g^0) < M + \mu$; and

(ii) $\Pr\{V[\psi(X,g^0)] - U[\psi(X,g^0)] > M + \varepsilon\} \geq \lambda$.

Let ε,λ be the values guaranteed by the Lemma. Let $\Gamma' = \{g | V(g) - U(g) < M + \varepsilon\}$. For any $g \in \Gamma'$, we have

$$\begin{aligned} E\{V[\psi(X,g)] - U[\psi(X,g)]\} = \; & \Pr\{\psi(X,g) \in \Gamma_1 \cup \Gamma_2\} \cdot \\ & E[V(\psi) - U(\psi) \mid \psi \in \Gamma_1 \cup \Gamma_2] + \Pr\{\psi(X,g) \in \Gamma_3 \backslash \Gamma'\} \cdot \\ & E[V(\psi) - U(\psi) \mid \psi \in \Gamma_3 \backslash \Gamma'] + \Pr\{\psi(X,g) \in \Gamma'\} \cdot \\ & E[V(\psi) - U(\psi) \mid \psi \in \Gamma']. \end{aligned}$$

Now $E[V(\psi) - U(\psi) | \psi \in \Gamma_1 \cup \Gamma_2] \geq 0$; $0 > E[V(\psi) - U(\psi) \mid \psi \in \Gamma_3 \backslash \Gamma'] \geq M + \varepsilon$; and $M + \varepsilon > E[V(\psi) - U(\psi) \mid \psi \in \Gamma'] \geq M$. Thus we can write

$$\begin{aligned} E\{V[\psi(X,g)] - U[\psi(X,g)]\} \geq \; & \Pr\{\psi(X,g) \notin \Gamma'\}(M + \varepsilon) + \\ & \Pr\{\psi(X,g) \in \Gamma'\} M = M + \varepsilon \Pr\{\psi(X,g) \notin \Gamma'\} \end{aligned}$$

for all $g \in \Gamma'$.

Let $\mu = \lambda\varepsilon$ for the λ and ε guaranteed by the Lemma, and let g^0 be as in the Lemma. Part (ii) of the Lemma guarantees

$$\Pr\{\psi(X,g) \notin \Gamma'\} \geq \lambda,$$

and since $g^0 \in \Gamma'$,

$$E\{V(\psi(X,g^0)) - U(\psi(X,g^0))\} \geq M + \varepsilon\Pr\{\psi(X,g^0) \notin \Gamma'\} \geq M + \lambda\varepsilon.$$

But by part (i) of the Lemma, $M + \lambda\varepsilon > V(g^0) - U(g^0)$. Thus

$$V(g^0) - U(g^0) < E\{V(\psi(X,g^0)) - U(\psi(X,g^0))\}.$$

Rearranging terms gives

$$\gamma_V(g^0) > \gamma_U(g^0)$$

which is the desired contradiction. Thus $\Gamma_3 = \emptyset$. By an identical argument, it can be shown that $\Gamma_2 = \emptyset$. Hence $V(g) = U(g)$ for all $g \in \Gamma$. □

<u>Proof of Lemma</u>: Suppose the Lemma does not hold true. Then for each ε and λ there is a $\mu_0 > 0$ such that

$$V(g) - U(g) < M + \mu_0 \Rightarrow \Pr\{V(\psi) - U(\psi) > M + \varepsilon\} < \lambda$$

where $\psi = \psi(X,g)$.

Let $\mu(\varepsilon,\lambda)$ be the largest such μ_0 for a given ε and λ. We will show that a state $g \in \Gamma$ can be found which makes EK_{δ_U} arbitrarily large; in particular; that for any P, $0 \leq P < 1$ and any k, there is a $g \in \Gamma$ for which $\Pr\{K_{\delta_U} \geq k\} \geq P$. Let $\lambda \geq P^{1/k}$ and let $\varepsilon_k = M$. Define a series $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ by $\varepsilon_{n-1} = \mu(\varepsilon_n, \lambda)$ for $n = 2,3,\ldots,k$. Let g^0 be such that $V(g^0) - U(g^0) < M + \mu(\varepsilon_1,\lambda)$; Such a g^0 exists by definition of M and because $\mu(\varepsilon_1,\lambda) > 0$. Then

$$\Pr\{V(\psi(X,g^0)) - U(\psi(X,g^0)) > M + \varepsilon_1\} < \lambda.$$

For a $g^1 = \psi(x,g^0)$ which <u>does</u> give $V(g^1) - U(g^1) < M + \varepsilon_1$, we have

$$\Pr\{V(\psi(X,g^1)) - U(\psi(X,g^1)) > M + \varepsilon_2\} < \lambda.$$

Thus with probability $\lambda^k \geq P$, each new state from g^1 through g^k will have

$$V(g^i) - U(g^i) < 0$$

so $g^i \in \Gamma_3$, and as was shown earlier in the proof of the Theorem, $\delta_U(g^i)$ = continue, for each of these g^i. Since EK_{δ_U} can be made arbitrarily large, so it is possible to get

$$U(g) = E\{v_0[\psi[X^1,\ldots,X^{K_{\delta_U}}), g] - cK_{\delta_U}\}$$

arbitrarily small (large negative). This contradicts the boundedness of U also demonstrated earlier in the proof of the Theorem. □

We now know that the functional equation for the voter's sequential search problem has exactly one solution $v^*(g)$. As shown above, this value is achieved by a sampling plan δ satisfying

$$\delta(g) = \begin{cases} \text{continue if } v_0(g) < Ev^*(\psi(X,g)) - c \\ \text{stop if } v_0(g) > Ev^*(\psi(X,g)) - c. \end{cases}$$

In addition we have seen that such a plan must terminate in finitely many steps with probability 1. In the subsequent sections, we shall use this well-defined functional equation to examine the nature of the optimal sampling plan.

3. OPTIMAL SAMPLING AND THE COST OF SEARCH

Intuitively, one would expect that in any search problem, the sampling plan would require less sampling as the cost of search increased. In most economics models of sequential search, this result obtains fairly easily, even for example in Rothschild's [1974] fairly general model of searching for the lowest price. However, the state space Γ for the voter's search problem is considerably more complicated than the state space for most of the sequential search models in the literature, where usually the state is described simply by the (scalar) value of the last observation. Thus the proof of this search cost result is no longer trivial. Also, there can be no "level" which describes the cutoff of sampling given a search cost, as in searching for prices, since here the state space is of not one but infinite dimension. Thus we seek to establish using Theorems 3 and 4 both whether and in what sense the amount of sampling responds negatively to increases in search cost. We first make the following convenient definition:

Definition: If g_1 and g_2 are priors such that

$$\inf_{y \text{ s.t. } g_1(y)>0} u(y) \geq \sup_{y \text{ s.t. } g_2(y)>0} u(y),$$

then $g = (g_1, g_2)$ is said to be an unambiguous prior.

Theorem 3: If g is an unambiguous prior then in the infinite horizon search problem, the optimal sampling plan will allow no observation to be taken if $c > 0$.

Proof: We first consider the one-period horizon problem, proceed by induction to the finite-horizon problem, and then taking limits obtain the general result.

For the 1-period horizon case, the functional equation is

$$v_1(g) = \max\{v_0(g),\ Ev_0(\psi(X,g)) - c\}.$$

In this case, sampling must stop if

$$Ev_0(\psi(X,g)) - v_0(g) < c.$$

(Notice that the left-hand expression does not depend on c.) Without loss of generality, assume $\underset{g}{E}u(Y_1) \geq \underset{g}{E}u(Y_2)$. Then also

$$\underset{\psi(x,g)}{E} u(Y_1) \geq \underset{\psi(x,g)}{E} u(Y_2)$$

for all x, since the support of $\psi(x,g)$ is contained in the support of g. Hence

$$\begin{aligned} \underset{f}{E}v_0(\psi(X,g)) &= \underset{f}{E}\ \underset{\psi(X,g)}{E} u(Y_1) \\ &= \underset{f}{E}[\int_S u(y_1)g(y|X)dy] \\ &= \underset{f}{E}\ \underset{g}{E}[u(Y_1)|X], \end{aligned}$$

a conditional expectation, which is just

$$= \underset{g}{E}\ u(Y_1) = v_0(g).$$

Since $\underset{f}{E}v_0(\psi(X,g)) - v_0(g) = 0$ for unambiguous g, there must never be any sampling if $c > 0$, although the individual should be indifferent between sampling or not sampling if $c = 0$.

In the general finite-horizon case, an optimal sampling plan says to stop observing whenever

$$Ev_{n-1}(\psi(X,g)) - v_0(g) < c$$

when n periods remain. Assuming that the one-period result applies to the n-period case, we show it holds also for the (n+1)-period problem. The assumption says

$$v_0(g) = Ev_{n-1}(\psi(X,g)).$$

Therefore the (n+1)-period problem is reduced to the one-period case because sampling may as well stop after one observation. Thus we know

$$v_0(g) = Ev_n(\psi(X,g)).$$

Finally, from the proof of Theorem 1 we have

$$Ev^*(\psi(X,g)) = \lim_{n\to\infty} Ev_n(\psi(X,g))$$

which by the above remarks

$$= v_0(g).$$

Thus there need be no sampling in the infinite-horizon problem if g is unambiguous. □

Theorem 4: If g_1 and g_2 are uniform, and f (here, $f = f_1 = f_2$) is not uniform on S, then there is a value $c > 0$ such that the optimal sampling plan will require that an observation be taken at $g = (g_1, g_2)$.

Proof: This proof proceeds generally as that of Theorem 3 from the one-period case to the general case.

Notice that

$$\underset{\psi(X,g)}{E} u(Y_1) \quad \text{and} \quad \underset{\psi(X,g)}{E} u(Y_2)$$

are measurable functions of X_1 and X_2, respectively; since X_1 and X_2 are independent, those two expectations are themselves independent random variables. Let Δ_1 and Δ_2 be their cumulative distribution functions.

In the same way,

$$\max\{\underset{\psi(X,g)}{E} u(Y_1), \underset{\psi(X,g)}{E} u(Y_2)\}$$

is another random variable; let H be its distribution function. H is then given by

$$\begin{aligned} H(p) &= \Pr \{\max[\underset{\psi}{E}u(Y_1), \underset{\psi}{E}u(Y_2)] \leq p\} \\ &= \Pr \{\underset{\psi}{E}u(Y_1) \leq p \text{ and } \underset{\psi}{E} u(Y_2) \leq p\} \\ &= \Delta_1(p)\Delta_2(p). \end{aligned}$$

Clearly if $\Delta_1(p) = \Delta_2(p) = 1$ then $H(p) = 1$, and if either $\Delta_i(p) = 0$ then $H(p) = 0$.

Since g_1 and g_2 are uniform on S and f is not, the new expectations $\underset{\psi}{E}u(Y_i)$ depend completely upon the observation X. Clearly if $X = (x_1,x_2)$ gives

$$\underset{\psi(x,g)}{E} u(Y_1) > \underset{\psi(x,g)}{E} u(Y_2)$$

then $X = (x_2,x_1)$ will give

$$\underset{\psi(x,g)}{E} u(Y_1) < \underset{\psi(x,g)}{E} u(Y_2).$$

If f has no mass points, then

$$\Pr\{\underset{\psi(X,g)}{E} u(Y_1) > \underset{\psi(X,g)}{E} u(Y_2)\} = 1/2.$$

In particular, $H(p) < \Delta_1(p)$ and $H(p) < \Delta_2(p)$ each occur on a set of positive probability, and for all p, $H(p) \leq \Delta_1(p)$ and $H(p) \leq \Delta_2(p)$. That is, H strictly stochastically dominates both Δ_1 and Δ_2. It is therefore known that

$$\underset{g}{E} u(Y_1) < \underset{f}{E} \max \{\underset{\psi}{E} u(Y_1), \underset{\psi}{E} u(Y_2)\}.$$

This is a consequence of the conditional probability property that $\underset{f}{E}\underset{\psi}{E} U(Y_1) = \underset{g}{E} u(Y_1)$ (see proof of Theorem 3) and of stochastic dominance. To see this, let B and $\overline{B}$ be the lower and upper bounds of the new random variables (recall that u is a bounded function). Using integration by parts,

$$\underset{f}{E}[\underset{\psi}{E} u(Y_1)] = \int_B^{\overline{B}} p \, d\Delta_1 = p \, \Delta_1(p) \Big|_B^{\overline{B}} - \int_B^{\overline{B}} \Delta_1(p)dp$$

$$= \overline{B} - \int_B^{\overline{B}} \Delta_1(p)dp,$$

and similarly

$$\underset{f}{E} \max \{\underset{\psi}{E}\, u(Y_1), \underset{\psi}{E}\, u(Y_2)\} = \int_{\underline{B}}^{\overline{B}} p dH$$

$$= \overline{B} - \int_{\underline{B}}^{\overline{B}} H(p)dp.$$

But since $\Delta_1(p) > H(p)$ on a set of positive measure and $\Delta_1(p) \geq H(p)$ everywhere,

$$\int_{\underline{B}}^{\overline{B}} \Delta_1 dp > \int_{\underline{B}}^{\overline{B}} H\, dp.$$

Hence

$$\int_{\underline{B}}^{\overline{B}} p\, d\Delta_1 < \int_{\underline{B}}^{\overline{B}} p\, dH,$$

that is

$$\underset{f}{E}\,\underset{\psi}{E}\, u(Y_1) < \underset{f}{E} \max \{\underset{\psi}{E}\, u(Y_1), \underset{\psi}{E}\, u(Y_2)\},$$

and similarly

$$\underset{f}{E}\,\underset{\psi}{E}\, u(Y_2) < \underset{f}{E} \max \{\underset{\psi}{E}\, u(Y_2), \underset{\psi}{E}\, u(Y_2)\}.$$

Therefore

$$v_0(g) < \underset{f}{E}\, v_0(\psi(X,g))$$

for uniform g. This holds for other g as well (for example whenever $g_1 = g_2$) but especially for uniform g since there the stochastic dominance is most pronounced.

In the finite-horizon case, assume the n-period result:

$$v_0(g) < \underset{f}{E} v_{n-1}(\psi(X,g)).$$

Now in general for dynamic programming problems, $v_n(g) \geq v_{n-1}(g)$ for all g: an extra period just affords the opportunity to take an extra observation if desirable (this is easily proved by induction on n). Hence

$$\underset{f}{E}\, v_n(\psi(X,g)) \geq \underset{f}{E}\, v_{n-1}(\psi(X,g)) > v_0(g).$$

Finally since $E\, v^*(\psi(X,g)) = \lim_{n\to\infty} E\, v_n(\psi(X,g))$, and the Ev_n are an increasing sequence,

$$E\, v^*(\psi(X,g)) > v_0(g).$$

Thus there is a c > 0 small enough so that an observation must be taken at g. □

We can now make use of Theorems 3 and 4 to explain in what sense a voter will search less if his search costs increase. Suppose first that f, g and u are continuous and that S is compact; then the value-of-search function

$$E\, v^*(\psi(X,g)) - v_0(g)$$

is a continuous function of g under an appropriate topology (for example, if $S \subseteq \mathbb{R}$ we could use the topology on the set of bounded functions on S defined by the supremum metric). Then if c is a value such that sampling occurs for some priors but not for others, and if $c' > c$, there will be a set of priors which require sampling at c but not at c'. Furthermore, this set will be substantial in the sense that it will include all density functions in some open set of the aforementioned topology. If discontinuities are present, of course, there may be ranges of c over which no change in sampling occurs. Thus if the voter faces a change in search costs, there will be a "large" number of priors at which he will change his sampling behavior, stopping to choose a candidate where he could previously have continued gathering information or vice versa. For some priors and some sequences of observations, in particular, the voter will cease gathering information earlier under a higher search cost; he may be forced to choose a candidate while there is less certainty about the candidates' positions.

4. OPTIMAL SAMPLING AND RISK

One of the primary characteristics of a search problem is of course the risk involved, since it is risk which makes sampling desirable in the first place. Intuitively, it seems that changes in risk should be related directly to changes in the desirability of sampling. In this section we shall attempt to explicate this relationship as far as is possible for the general search problem.

It was seen in Theorem 4 that the value of an additional observation at state g, apart from search cost, is

$$Ev^*(\psi(X,g)) - v_0(g),$$

and that this quantity varies with the nature of g. Without many additional assumptions on u, f, and g it is impossible to examine directly the relationships between risk in f and g, risk preference in u, and optimal search strategies. It is possible, however, to link riskiness of the derived distributions in Theorem 4, namely Δ_1 and Δ_2, to search behavior. In particular, it can be shown that an increase in the risk of the random variable $\underset{\psi(X,g)}{E} u(Y_i)$ for i=1 or 2 will increase the value of an additional observation at state g.

Because u is a bounded function, $\underset{\psi(X,g)}{E} u(Y_i)$ is a random variable whose values lie in a bounded interval of the real line. Rothschild and Stiglitz [1970] have shown that at least for random variables of this type, the notion of increasing risk is captured precisely by a mean-preserving spread in the distribution function Δ_i. If Δ_i is the original distribution and Δ_i^* a more risky distribution, Δ_i and Δ_i^* are said to differ by a mean-preserving spread if

(i) the mean is preserved, i.e. $\int p d\Delta_i = \int p d\Delta_i^*$;

(ii) $\int \Delta_i^*(p) - \Delta_i(p) dp = 0$ (so that Δ_i^* and Δ_i are both proper distribution functions); and

(iii) there is a value z in the range of $\underset{\psi(X,g)}{E} u(Y_i)$ such that

$$\Delta_i^*(p) - \Delta_i(p) = \begin{cases} \geq 0 \text{ for } p \leq z \\ \leq 0 \text{ for } p > z. \end{cases}$$

The latter just defines $\Delta_i^* - \Delta_i$ as in increase in risk rather than as a decrease or some mixture thereof. (For an explanation of these properties, see Rothschild and Stiglitz [1970].)

Suppose that one of the Δ_i, say Δ_1, undergoes a mean-preserving spread to Δ_1^*. Since $\underset{\psi(X,g)}{E} u(Y_1)$ is just a transformation of the random variable X, Δ_1 is itself a function of the distribution of X which is in turn defined by the prior g(y) and the conditional distribution $f(x|y)$: the density of X is

$$f(x) = \int g(t) f(x|t) dt.$$

Notice, then, that the mean-preserving spread $\Delta_1^* - \Delta_1$ is the result of some change $f^* - f$ in the marginal density of X (which may not be a mean-preserving spread). It will be necessary here to make an additional assumption about the change underlying $\Delta_1^* - \Delta_1$, namely that there is an equivalent change both in g and in all its possible updates $\psi(x,g)$. That is, not only is there a mean-preserving spread in $\underset{\psi(X,g)}{E} u(Y_1)$ but also in every $\underset{\psi(X,h)}{E} u(Y_1)$ where $h = \psi(x,g)$ for some $x \in S \times S$. Substantively, this might be thought of as resulting from some exogenous factor which shifts all hypothetical priors to a state of greater uncertainty; or simply as an assumption that the change in g affects all its updates equivalently. Formally, it is a necessary assumption in order to get from the 1-period to the n-period horizon in the following proof.

Finally, recall that in Theorem 4 the value of an additional observation in the 1-period horizon problem was shown to be

$$\int p dH - \int p d\Delta_1$$

where $H(p) = \Delta_1(p)\Delta_2(p)$ is the distribution of $v_0(\psi(X,g))$ and where it is assumed $v_0(g) = \underset{g}{E}\, u(Y_1)$. The value of search was shown to be positive by showing that both $\int p dH - \int p d\Delta_1$ and $\int p dH - \int p d\Delta_2$ are positive. With these preliminaries in mind, we can now prove the following theorem relating risk to the value of search:

Theorem 5: Suppose that g_1 undergoes a transformation of the type described above, so that $\underset{\psi(X,h)}{E} u(Y_1)$ undergoes a mean-preserving spread for $h = g$ and all $h = \psi(X,g)$. Then the value of an additional observation at state g is (weakly) increased.

Proof: The proof proceeds by induction, using a limiting argument for the infinite-horizon problem.

We wish to show that, for i = 1 or 2,

$$\underset{f}{E}\, v_0(\psi(X,g)) - \underset{g}{E}\, u(Y_i)$$

is increased. If H_0 is the distribution function of $v_0(\psi(X,g))$, this difference can be written

$$\int p dH_0 - \int p d\Delta_i = \int \Delta_i(p) - H_0(p) dp,$$

the latter using integration by parts. If Δ_1^* and $H_0^* = \Delta_1^*\Delta_2$ are the new distributions, the difference made by the mean-preserving spread is, for i = 1,

$$\int \Delta_1^*(p) - H_0^*(p)dp - \int \Delta_1(p) - H_0(p)dp$$
$$= \int \Delta_1^*(p) - \Delta_1(p)dp + \int H_0(p) - H_0^*(p)dp.$$

By property (ii) of mean-preserving spreads, the first term is 0, so we have

$$= \int H_0(p) - H_0^*(p)dp = \int \Delta_2(p)[\Delta_1(p) - \Delta_1^*(p)]dp$$
$$= -\int \Delta_2(p)\ [\Delta_1^*(p) - \Delta_1(p)]dp.$$

Since $\Delta_2(p)$ is an increasing function, properties (ii) and (iii) imply that

$$\int \Delta_2(p)\ [\Delta_1^*(p) - \Delta_1(p)]dp < 0.$$

Similarly, $\int H_0(p) - H_0^*(p)dp > 0$ implies

$$\int \Delta_2(p) - H_0^*(p)dp - \int \Delta_2(p) - H_0(p)dp > 0.$$

Thus for the 1-period problem, the value of search is increased.

Assume now that this result also holds for the n-period problem in the same way: that

$$\int H_{n-1}(p) - H_{n-1}^*(p)dp > 0$$

where H_{n-1} is the distribution function of $v_{n-1}(\psi(X,g))$. We wish to prove these inequalities for H_n. Now

$$\int pdH_{n-1} = \int v_{n-1}(\psi(x,g))f(x)dx$$

so the induction assumption can be rewritten as

(2) $$\int v_{n-1}(\psi(x,g))[f^*(x) - f(x)]dx \geq 0.$$

Then

(3) $$\int v_n(\psi(x,g))[f^*(x) - f(x)]dx$$

$$= \int \max \{v_0(\psi(x,g)), \underset{W}{E}\, v_{n-1}[\psi(W,\psi(x,g))] - c\} [f^*(x) - f(x)]dx,$$

using the functional equation for the n-period problem. Since v_0 and v_{n-1} are both bounded below, this is

(4) $$\geq \int v_0(\psi(x,g))[f^*(x) - f(x)]dx$$

which is nonnegative as shown above (the 1-period problem); likewise,

$$\int v_n[f^* - f]dx \geq \int [\underset{W}{E}\, v_{n-1}[\psi(W,\psi(x,g))] - c][f^*(x) - f(x)]dx$$

$$= \int \underset{W}{E} v_{n-1}[\psi(W,\psi(x,g))][f^*(x) - f(x)]dx$$

since f* and f are both density functions and integrate to 1. Using Bayes' formula, $\psi(W,\psi(x,g)) = \psi(x,\psi(W,g))$. Since the integral over dx and that in $\underset{W}{E}\, v_{n-1}$ are over the entire range of the random variables, we can reverse the order of integration to get

$$= \underset{W}{E} \int v_{n-1}[\psi(x,\psi(W,g))][f^*(x) - f(x)]dx.$$

But now because of the hypothesized changes in $\psi(W,g)$ for each value of W, the induction assumption (2) applies to the inside integral for all values of W, and hence the expectation is non-negative: thus

$$(5) \quad \int \{\underset{W}{E} v_{n-1}[\psi(W,\psi(x,g))] - c\}[f^*(x) - f(x)]dx \geq 0.$$

Equation (3) together with inequalities (4) and (5) then give the desired result for the (n+1)-period problem.

Finally, using the same limiting argument as was used in Theorem 1 to define $v^* = \lim_{n\to\infty} v_n$, we have

$$\int H(p) - H^*(p)dp \geq 0$$

where H is the distribution function of $v^*(\psi(X,g))$. Thus for the general voter's search problem a mean-preserving spread of the kind hypothesized yields an increase (or no change) in the value of the next observation,

$$\underset{f}{E}\ v^*(\psi(X,g)) - v_0(g).$$

□

Because a mean-preserving spread in f or g does not necessarily cause a mean-preserving spread in Δ_i, we cannot directly relate changes in risk in the voter's basic subjective distributions to changes in sampling behavior. But Theorem 5 is still meaningful as a statement about risk changes, since it does deal with uncertainties in the voter's future expected utility. It should be noted that since it is a utility value and not an issue position which is the random variable in question, the risk preferences of the voter, that is, the shape of his utility function over issue positions, does not affect the result in Theorem 5. We have shown in general, then, that as the expected utility of either candidate after the next observation becomes more risky, the value of search and thus the amount of information that will be gathered decrease, in the same sense as in Section 3.

5. OBSERVATIONS AND CHOICE OF CANDIDATES

The objective of the individual who is engaging in a sequential sampling process, is, of course, to choose between the alternatives offered. In the present model, that choice is the actual political behavior in question. It is useful, then, to examine what relationship exists between the information gathered and the choice of candidates which is finally made.

In the first place, one might expect that if an observation says candidate 1 is better than candidate 2, candidate 1 will not become less desirable in the observer's estimation relative to candidate 2. For a particularly simple combination of preferences and expectations, Theorem 6 shows that this is indeed the case. An important observable implication is that such an individual will never observe that candidate 1 is better and immediately stop and choose candidate 2; in other words, an individual meeting the requirements of Theorem 6 who chooses candidate 1 must have last observed candidate 1 to be preferred.

In the theorem which follows, α represents the individual's most preferred policy position and f is the sampling distribution of x_i given y_i for i = 1 or 2.

<u>Theorem 6</u>: Suppose g_1 and g_2 are uniform. Suppose $f(x_i|s) = f^*(\|x_i - s\|)$ and $u(s) = u^*(\|\alpha - s\|)$ for decreasing functions f* and u*, that is, f and u have circular contours. If x is observed such that $u(x_1) > u(x_2)$ and such that the supports of $f(x_1|\cdot)$ and $f(x_2|\cdot)$ lie within S, then

$$\underset{\psi(x,g)}{E} u(Y_1) > \underset{\psi(x,g)}{E} u(Y_2).$$

Proof: The strategy of this proof is to show that

$$(6) \quad \int_B u(y_1)g_1(y_1|x_1)dy_1 > \int_B u(y_2)g_2(y_2|x_2)dy_2$$

for every ball B centered at α; then taking B large enough so $\int_{S-B} g(y|x)dy$ is arbitrarily small, the desired inequality follows. Let H be the hyperplane in $\mathbb{R}^n$ (where $S \subset \mathbb{R}^n$) which bisects the line segment from x_1 to x_2. This divides S into two open half-spaces, $H_1 = \{s | \|s - x_1\| < \|s - x_2\|\}$ and $H_2 = \{s | \|s - x_1\| > \|s - x_2\|\}$. Note that $\alpha \in H_1$.

If $B \subset H_1$, then (6) follows easily since $\|s - x_1\| < \|s - x_2\|$ $\forall s \in B \Rightarrow f(x_1|s) > f(x_2|s)$ on B and since g_i uniform means that $g_i(s|x_i) = f(x_i|s)$. For larger B, the inequality is less obvious, and several more definitions are necessary. Let R be the function which assigns to each $s \in S$ its reflection across H. Let P be the function which assigns to every s its projection on H. Notice that $P(s) = P(R(s))$ for every s, that $R(x_1) = x_2$, and that $R(R(s)) = s$ for every s.

We must first prove that for every s, $f(x_1|s) = f(x_2|R(s))$. Consider the triangle with vertices x_1, s, and t_1 where

$$t_1 = x_1 + (P(x_1) - x_1) + (s - P(s)) = P(x) + s - P(s),$$

and the triangle with vertices x_2, $R(s)$ and t_2 where

$$t_2 = x_2 + (P(x_2) - x_2) + (R(s) - P(R(s)))$$
$$= P(x_2) + R(s) - P(R(s))$$
$$= P(x_1) - R(s) - P(s).$$

The angle of the first triangle at t_1 is a right angle:

$$(x_1 - t_1) \cdot (t_1 - s) = (x_1 - P(x_1) + s - P(s)) \cdot (P(x_1) - P(s))$$
$$= [x_1 - P(x_1)] \cdot [P(x_1) - P(s)] + [s - P(s)] \cdot [P(x_1 - P(s)].$$

$x_1 - P(x_1)$ and $s - P(s)$ are perpendicular to H; and since $P(x_1) \in H$ and $P(s) \in H$, $P(x_1) - P(s)$ is parallel to H. Therefore both terms are zero, as required.

Likewise the angle of the second triangle at t_2 is a right angle:

$$(R(x_1) - t_2) \cdot (t_2 - R(s)) = [R(x_1) - P(x_1) + R(s) - P(s)] \cdot [P(x_1) - P(s)]$$
$$= [R(x_1) - P(x_1)] \cdot [P(x_1) - P(s)] + [R(s) - P(s)] \cdot [P(x_1) - P(s)]$$
$$= 0.$$

Furthermore, the two sides adjacent to the right angles are of equal lengths in the two triangles:

$$(t_1 - s) = P(x_1) + s - P(s) - s = P(x_1) - P(s)$$

$$(t_2 - R(s)) = P(x_1) + R(s) - P(s) - R(s) = P(x_1) - P(s).$$

As for the other sides,

$$(x_1 - t_1) = x_1 - P(x_1) + s - P(s)$$

while

$$(R(x_1) - t_2) = R(x_1) - P(x_1) + R(s) - P(s).$$

By definition of R and P,

$$P(x_1) = (1/2)x_1 + (1/2)R(x_1),$$

so

$$x_1 - P(x_1) = -[R(x_1) - P(x_1)].$$

Likewise,

$$s - P(s) = -[R(s) - P(s)].$$

Thus,

$$\|x_1 - t_1\| = \|x_1 - P(x_1) + s - P(s)\| = \|-[R(x_1) - P(x_1)] - [R(s) - P(s)]\| = \|R(x_1) - P(x_1) + R(s) - P(s)\| = \|R(x_1) - t_2\|$$

as required.

Since the two right triangles have legs of equal length, their hypotenuses are of equal length:

$$\|x_1 - s\| = \|R(x_1) - R(s)\|$$

so under our assumption about the form of f,

$$f(x_1|s) = f(x_2|R(s)).$$

For convenience, let $B_1 = B \cap H_1$ and $B_2 = B \cap H_2$. We must next show that for all $s \in B_2$, $u(s) < u(R(s))$. For any s, the definition of R means that H is the perpendicular bisector of the line segment between s and R(s). In particular, this can be applied to α and $R(\alpha)$. Now since $B_2 \subset H_2$, for any $s \in B_2$ we have

$$\|R(\alpha) - s\| < \|\alpha - s\|.$$

As shown above, though, $\|R(\alpha) - s\| = \|R(R(\alpha)) - R(s)\| = \|\alpha - R(s)\|$. Thus since we assumed u to have circular contours,

$$s \in B_2 \Rightarrow \|\alpha - R(s)\| < \|\alpha - s\| \Rightarrow u(R(s)) > u(s)$$

as required.

We can divide B into two parts: $B_2 \cup R(B_2)$ and $B_1 \setminus R(B_2)$. (Notice that $R(B_2) \subset B$ since as just shown for $s \in B_2$, $\|\alpha - R(s)\| < \|\alpha - s\|$, which in turn is less than the radius of B.) On the first part of B,

$$\int_{B_2 \cup R(B_2)} u(s)[g_1(s|x_1) - g_2(s|x_2)]ds$$

$$= \int_{B_2 \cup R(B_2)} u(s)[f(x_1|s) - f(x_2|s)]ds$$

since g is uniform over S and since S contains the supports of $f(x_1|\cdot)$ and $f(x_2|\cdot)$. Separating terms,

$$= \int_{B_2} u(s)f(x_1|s)ds - \int_{B_2} u(s)f(x_2|s)ds + \int_{R(B_2)} u(s)f(x_1|s)ds$$

$$- \int_{R(B_2)} u(s)f(x_2|s)ds$$

$$= \int_{B_2} u(s)f(x_1|s)ds - \int_{B_2} u(s)f(x_2|s)ds + \int_{B_2} u(R(s))f(x_1|R(s))ds$$

$$- \int_{B_2} u(R(s))f(x_2|R(S))ds$$

$$= \int_{B_2} u(s)f(x_1|s)ds - \int_{B_2} u(s)f(x_2|s)ds + \int_{B_2} u(R(s))f(x_2|s)ds$$

$$- \int_{B_2} u(R(s))f(x_1|s)ds,$$

the latter using the fact that $f(x_1|s) = f(x_2|R(s))$. Since on B_2, $u(s) < u(R(s))$, this is

$$> \int_{B_2} u(s)[f(x_1|s) - f(x_2|s) + f(x_2|s) - f(x_1|s)]ds = 0.$$

For the second part of B, notice that $s \in B_1' \setminus R(B_2) \subset H_1$ implies that

$$f(x_2|s) < f(x_2|R(s)),$$

so

$$\int_{B_1 \setminus R(B_2)} u(s)g(s|x_2)ds < \int_{B_1 \setminus R(B_2)} u(s)g(R(s)|x_2)ds.$$

But

$$g(R(s)|x_2) = f(x_2|R(s)) = f(x_1|s) = g(s|x_1),$$

so

$$\int_{B_1 \setminus R(B_2)} u(s)g(s\ x_2)ds < \int_{B_1 \setminus R(B_2)} u(s)g(s)|x_2)ds.$$

We are now done, for

$$\begin{aligned}\int_B u(s)g_1(s|x_1)ds &= \int_{B_2 \cup R(B_2)} u(s)g_1(s|x_1)ds \\ &\quad + \int_{B_1 \setminus R(B_2)} u(s)g_1(s|x_1)ds \\ &> \int_{B_2 \cup R(B_2)} u(s)g_2(s|x_2)ds \\ &\quad + \int_{B_1 \setminus R(B_2)} u(s)g_2(s|x_2)ds \\ &= \int_B u(s)g_2(s|x_2)ds.\end{aligned}$$

Since the supports of g_1, g_2, and f are within S, letting B be arbitrarily large then gives

$$\int_S u(s)g_1(s|x_1)ds > \int_S u(s)g_2(s|x_2)ds$$

or

$$\underset{\psi}{E}\, u(Y_1) > \underset{\psi}{E}\, u(Y_2). \qquad \square$$

The conditions of Theorem 6 are unfortunately very strong. The individual is required to have no real idea before sampling of what the candidates' positions are; to have a sampling distribution with circular contours; and to have Type I preferences. Any of these alone could be finessed by transforming the issue space, provided the issues which define S are independent of each other with respect to g, f, or u. However, all three conditions must hold at the same time, that is, for the same such transformation of S. Also, the observation (x_1,x_2) is required to be sufficiently removed from the boundary of S that $f(x|\cdot)$ is zero outside S. In the Appendix to this chapter, we will examine a series of counterexamples demonstrating the importance of these conditions. As it turns out, relaxing any single condition of Theorem 6 may cause the theorem to fail. None of the counterexamples uses preferences or subjective distributions which are any more pathological or unusual than those required for Theorem 6.

The counterexamples in the Appendix are important for another reason as well. In showing several ways in which small differences in utility functions, priors, and sampling distributions can bring about opposite results in behavior, they demonstrate the danger of making casual assertions about how uncertainty will affect voting. Because of this danger, it is not sufficient in modelling the electoral process to design models using complete information and then assume that uncertainty will provide whatever continuities or irregularities are necessary to explain observed phenomena not in keeping with the simpler model. Instead, studies like the present one are required in order to analyze behavior under uncertainty and state exactly what special conditions on preferences, priors, and so on will lead to what kinds of sampling and voting behavior.

6. ELECTIONS WITH MORE THAN TWO CANDIDATES

The results of Theorems 1 - 6 can be easily extended to the case where the voter must choose from among more than two candidates. We simply let $Y = (Y_1, Y_2, \ldots, Y_n)$ and

$$v_0(g) = \max \{E_g u(Y_1), \ldots, E_g u(Y_n)\}$$

where n is the number of candidates, and proceed as before. The presumption that u depends only on the chosen candidate's position is less tenable in this setting, since it may matter to some voters whether they are voting for a candidate who has any hope of winning. However, this can easily be incorporated into the model by letting each candidate's likely performance in the election be just another "issue" defining S_1, and treating it accordingly.

Theorems 1 and 2, on existence and uniqueness of the functional equation, go through exactly as given. For Theorem 3, an unambiguous prior may simply be taken to be one for which

$$\inf_{y \text{ s.t. } g_1(y)>0} u(y) \geq \sup_{y \text{ s.t. } g_i(y)>0} u(y)$$

for $i = 2,\ldots,n$. This establishes that there is still some large class of priors for which no sampling is ever necessary, as required for the result on search costs and sampling. Theorems 4 and 5 generalize in the obvious manner, letting $H_0(p) = \Delta_1(p)\Delta_2(p) \ldots\Delta_n(p)$ in the proof of stochastic dominance. Finally, Theorem 6 and the ensuing counterexamples apply to any pair of candidates.

7. NONVOTING

By describing the "voter's" problem as one of choosing a preferred candidate, we have ignored the question of whether or not the individual votes at all. This question cannot be addressed simply by imposing a comparison between $v_0(\psi)$ and the cost of voting after sampling is completed, because the possibility of abstaining should have an effect on the sampling process itself.

Thus the cost of voting and the option of nonvoting must be incorporated into the search problem from the beginning. As it turns out, it is easy to incorporate "abstention due to alienation" (Hinich and Ordeshook [1969]) into the model.

Let c_v represent the cost of voting, again in terms of the utility u**(z,L) of foregone leisure or other private goods. Let u_0 represent the utility of not voting relative to the utility u(y) of voting for a candiate at y. Then we can generalize the search model simply by defining

$$v_0(g) = \max \{\underset{g}{E}\, u(Y_1) - c_v, \underset{g}{E}\, u(Y_2) - c_v, u_0\}.$$

This closely resembles the three-candidate election dealt with in the last section; in fact, the only difference lies in some interpretations of terms. Again Theorems 1 and 2 go through without modification. In Theorem 3, it is now necessary to define a class of unambiguous priors with respect to u_0; this may be done by treating nonvoting exactly like a third candidate (about whom there is no uncertainty). In Theorem 4, u_0 is a third "random variable" with distribution Δ_3, where

$$\Delta_3(p) = \begin{cases} 0 \text{ if } p < u_0 \\ 1 \text{ if } p \geq u_0, \end{cases}$$

so $H(p) = 0$ for $p < u_0$ and $H(p) = \Delta_1(p)\Delta_2(p)$ for $p \geq u_0$. Letting $\underset{g}{E}\, u(Y_i) - c_v$ take the place of $\underset{g}{E}\, u(Y_i)$, the stochastic dominance result still goes through. There is one hitch, however: if c_v is very high or u_0 very high relative to the possible values of $u(Y_i)$, there might be no sampling even for uniform priors. In this case, the individual is certain to be a nonvoter and the search problem is irrelevant anyway. Finally, Theorem 6 and its counterexamples still hold; now, however, an individual may observe that candidate 1 is preferred and choose nonvoting.

The possibility of "abstention due to indifference" requires additional modifications in the model. Rather than letting u depend only on the position of the chosen candidate, we must allow it to depend also upon the difference in utility between the platform of the candidate chosen and that of his opponent; for example, we could consider $u(y_1,y_2) = \hat{u}\,(\max\{y_1,y_2\}, |y_1 - y_2|)$. For convenience we shall let $u(y_1,y_2)$ represent the utility of voting for the candidate whose true position is y_1 when the other candidate's true position is y_2. Theorems 1 and 2 are again unchanged, provided that u remains bounded. For Theorem 3, a major modification of the definition of unambiguous priors is required. For example, we might now say g is unambiguous if the original condition is met and if in addition $u(Y_1) - u(Y_2)$ is guaranteed large enough so that $u(y_1,y_2) - c_v > u_0$ with probability

1. Many such priors will exist if c_v and u_0 are small enough, but, depending on how $|y_1 - y_2|$ affects $u(y_1,y_2)$, u_0 and c_v could be small enough that voting and sampling may occur and yet too large for <u>any</u> "unambiguous" priors to exist. A different strategy of proof would be necessary, and might depend on exactly what form $u(y_1,y_2)$ takes. Still, it seems intuitively likely that higher search cost should under no circumstances lead to more sampling.

Theorem 4's proof is also inappropriate for the new formulation, since $\underset{\psi}{E}\, u(Y_1,Y_2)$ and $\underset{\psi}{E}\, u(Y_2,Y_1)$ both depend upon X_1 <u>and</u> X_2 and hence are no longer independent random variables; thus it is not true that $H(p) = \Delta_1(p)\Delta_2(p)$. But if we make further assumptions about the form of u, the problem might still be tractable. For example, suppose u is scaled so that $u(y_1,y_2) = -u(y_2,y_1)$. Then $\Delta_2(p) = 1 - \Delta_1(-p)$, and

$$H(p) = \Pr\{\max[E\,u(Y_1,Y_2),\ E\,u(Y_2,Y_1)] < p\}$$

$$= \Pr\{E\,u(Y_1,Y_2) < p \text{ and } E\,u\,(Y_2,Y_1) < p\}$$

$$= \Pr\{E\,u(Y_1,Y_2) < p\}\ \Pr\{E\,u(Y_2,Y_1) < p \mid E\,u(Y_1,Y_2) < p\}$$

$$= \frac{\Delta_1(p)\,[\Delta_1(p) - \Delta_1(-p)]}{\Delta_1(p)}$$

$$H(p) = \Delta_1(p) - \Delta_1(-p)$$

so that again $H(p) \leq \Delta_1(p)$ for all p and strictly so for some p; likewise for Δ_2. Thus stochastic dominance again holds, and Theorem 4 goes through. A more general possibility might be to simply require that

$$u(y_1,y_2) \geq u(y_1^*,y_2^*) \iff u(y_2,y_1) \leq u(y_2^*,y_1^*)$$

for all y_1, y_2, y^*_1, $y_2^* \in S$. In Theorem 5, the step in which $\int H_0 - H_0^* \, dp$ is shown to be positive is no longer valid using the new expression for $H(p) = H_0(p)$ given above. Finally, Theorem 6 does not seem to have a direct translation to this model, since the meaning of Type I preferences is no longer obvious at all, although again particular forms of u might prove tractable.

It should be noticed that the suggested models incorporating nonvoting are really broader than they might appear. They (especially the second one) can be specialized to any model in which voting behavior depends upon candidate positions and _not_ upon the probability of the individual's vote affecting the outcome of the election. Clearly, these models apply to voters for whom voting is an act of _consumption_ of candidates' issue positions (not simply the act of voting itself). But they also apply to _instrumental_ orientations in which an individual recognizes the possibility of affecting the outcome of the election but does not assign a probability to that event. An example of the latter kind of model is the minimax regret decision rule applied by Ferejohn and Fiorina [1974]. In another paper, Fiorina [1976] has shown that individuals' predictions of the closeness of an election do not have an independent effect on whether they vote, whereas perceived issue positions do. This empirical finding indicates that the decision to vote or abstain behaves generally as the other models in the present study would predict.

8. OTHER EXTENSIONS

Although the search models described here require that search costs and sampling distributions be fixed, and that observations come in pairs, we can use the models to gain some insight into voter behavior when those conditions are relaxed. Theorem 3 shows that if at state g an observation from sampling distribution f at cost c is taken, then an observation from f at a cost less than c will also be taken. Stated another way, if a voter is not presently sampling candidates' messages, i.e., paying attention to the campaign, it is possible that an observation of low enough cost will later come along which the individual will accept and evaluate. In fact for fixed f we can state the following corollary to Theorems 2 and 3:

Corollary: Let K(c) be the random variable which gives the number of observations which will be taken starting at state g under an optimal sampling plan, if search cost is c. Then K is a decreasing function of c for $c \geq 0$.

Proof: In the proof of Theorem 2 it was shown that $K(c) < \infty$ with probability 1. In Theorem 3 it was shown that increasing c causes some states to no longer require additional observations, while all others remain unchanged. Thus as possible updates of g fall into the former class, K(c) decreases as c increases. □

As for sampling distribution, it is clear that if $f(\cdot|y)$ is independent of y then sampling is worth nothing since $\psi(x,g) = g$ for all x. On the other hand, in the proof of Theorem 4 it was seen that for some g any other sampling distribution does yield a positive value of taking an observation. In addition to depending on g, however, the value of sampling for various f also depends on u; for example, it depends heavily upon the individual's risk preference. Any generalizations of the effects of changing f for the general model remain obscure.

Finally, relaxing the requirement that x_1 and x_2 be observed together changes the whole nature of the problem, adding to the choice of sampling or not sampling a choice of which candidate to observe. However, in Theorems 1 and 2 it was not required that f_1 be equal to f_2; by letting one of those be uniform on S for all y, we can achieve the effect of obtaining information about one candidate but not about the other. The results of those two theorems still apply in the following sense: for any given f_1 and f_2, one of which may be uniform, and any given search cost, the functional equation is well-defined and sampling will terminate finitely with probability 1. Theorems 3, 4, and 5 can also be reinterpreted along these lines.

In summary, a complete generalization of the present model would require giving the individual a choice between not sampling and sampling from different distributions (f_1, f_2) at different costs. Although specific results are not obtained immediately, it is apparent that the general structure of this sequential search problem, as represented by Theorems 1 - 6, would be retained.

Appendix

COUNTEREXAMPLES FOR THEOREM 6

In the following counterexamples, α represents the voter's bliss point; x_1 and x_2 are the observations of the two candidates; S is the issue space; and g and f are the prior and sampling densities.

Counterexample 1: In this example, all conditions of Theorem 6 are met except that S does not contain the support of $f(x_2|\cdot)$.

Let S be the interval [-20,10] in $\mathbb{R}$. Let $g_1 = g_2 = 1/30$ on S (and 0 elsewhere). Suppose $f(x_i|s) = 1/30$ on $[s - 15, s + 15]$. Let $\alpha = 0$ and $u(s) = -|\alpha - s|$. Finally, suppose $x_1 = -5$ and $x_2 = 10$. In general, if $g = 1/(b-a)$ on [a,b] and $f(x|y) = 1/(c-d)$ on $[x - d, x + c]$, then

$$g(y|x) = \frac{1/[(b-a)(c-d)]}{\int_{S\cap[x-d,x+c]} 1/[(b-a)(c-d)]dt + \int_{S\setminus[x-d,x+c]} 1/(b-a) \cdot 0 \, dt}$$

$$= \frac{1}{\int_{S\cap[x-d,x+c]} dt}.$$

In this example, then, $g(s|x_1) = 1/30$ on S and zero elsewhere, while $g(s|x_2) = 1/15$ on [-5,10] and zero elsewhere. Thus the posterior expectations in the example are

$$\underset{\psi(x,g)}{E} u(Y_1) = \int_{-20}^{10} -|s| \cdot 1/30 \, ds$$

$$= -2/3 \cdot 10 - 1/3 \cdot 5 = -25/3$$

and

$$\underset{\psi(x,g)}{E} u(Y_2) = \int_{-5}^{10} -|s| \cdot 1/15 \, ds$$

$$= -1/3 \cdot 5/2 - 2/3 \cdot 5 = -25/6$$

so

$$E\, u(Y_2) > E\, u(Y_1).$$

Notice that in this example $u(x_1) > \underset{g_i}{E}\, u(Y_i) > u(x_2)$ as well; such a strengthening of the condition $u(x_1) > u(x_2)$ would not eliminate this counterexample. □

Counterexample 1 is driven by the placing of x_2 near the boundary of S, and letting the uniform distributions take over. A slight weakening of the latter condition, however, makes the boundaries of S irrelevant, as is seen in the next example.

Counterexample 2: In this example, S does contain the supports of $f(x_i|s)$ for both i; but here although $g_1 = g_2$, the priors are not uniform. Let $S = \mathbb{R}$. The example would still work for S very large relative to the precision of g and f but still bounded; but the updating process, although approximately like that here, would be more complicated if the densities were truncated to S and normalized by adding a small constant. Thus for simplicity we let S be unbounded. Suppose $x_1 < \alpha < x_2$ with $x_2 - \alpha = (3/2)(\alpha - x_1)$. Let g_1 and g_2 be normal densities with mean $(1/2)(\alpha + x_1)$ and precision τ (the precision is the inverse of the variance). Let $f(x|s)$ be normal with mean s and precision τ. Let $u(s) = -|\alpha - s|$.

Thus to begin with $\underset{g}{E}\, u(Y_1) = \underset{g}{E}\, u(Y_2)$, and $u(x_1) > u(x_2)$. In general, if the prior has mean μ and precision τ and an observation x is taken from a sampling distribution with precision r, the posterior density is normal with mean $\tau\mu{+}rx/\tau{+}r$ and precision $\tau + r$ (see DeGroot [1970]). Here, then, $\psi(x_1, g_1)$ is normal with mean $(3/4)x_1 + (1/4)\alpha$ and precision 2τ; and $\psi(x_2, g_2)$ is normal with mean $(2/3)\alpha + (1/3)x_2$ and precision 2τ. Notice that

$$\alpha - (3/4)x_1 - (1/4)\alpha = (3/4)(\alpha - x_1)$$

while

$$(2/3)\alpha + (1/3)x_2 - \alpha = (1/3)(x_2 - \alpha) = (1/2)(\alpha - x_1).$$

Without computing the expectations, it is easy to see that since the mean of Y_2 is closer to α than is that of Y_1, and since both densities still have the same precision, the mean distance of Y_2

from α will be less than that of Y_1. Hence

$$\underset{\psi}{E}\, u(Y_2) > \underset{\psi}{E}\, u(Y_1). \qquad \square$$

In Counterexample 2, the problem is that although candidate 2 is inferior to candidate 1 both by prior expectation and for the observation, the bliss point lies <u>between</u> x_2 and candidate 2's prior expected position. The updating process then brings 2's updated expected position very near the bliss point. This possibility really represents the most important problem with applying the conclusion of Theorem 6 more generally: a position may have a given utility value but lie in <u>any</u> direction from the bliss point. Bayesian updating requires that the sampling distribution be fixed; for the case of a normal f, as in Counterexample 2, this means that observations wildly inconsistent with prior expectations still result in a posterior distribution with a <u>smaller</u> variance than that of the prior. The result, as we have seen, is that expectations of low utility plus observation of a low-utility position can result in a high-utility expectation. Such problems are unavoidable in any model in which a rational individual observes actual candidate positions and forms subjective likelihoods about them.

In the next two examples, we see that Type I preferences are also (loosely speaking) a necessity for Theorem 6.

Counterexample 3: In this example, all the conditions of Theorem 6 are met except that preferences are not Type I, that is, indifference curves are not circular. The example uses a discontinuous utility function whose indifference curve has flats and kinks, as well as a discontinuous sampling density. It will be indicated how the same idea can be extended to differentiable u and continuous f. Let S be the rectangle in $\mathbb{R}^2$ bounded by 0 and 11 on the horizontal axis (z_1) and by -1 and 11 on the vertical axis (z_2). Let A be all points in the triangle bounded by the lines $z_2 = 0$, $z_1 = 10$, and $z_1 = z_2$; assume A contains the two latter boundaries but not the first. Let $B = S \setminus A$. Let g_1 and g_2 be uniform on S. Let

$$u(s) = \begin{cases} 66/50 \text{ for } s \in A \\ 0 \text{ elsewhere.} \end{cases}$$

Let

$$f(x|s) = \begin{cases} 1/\pi \text{ for } \quad x-s \quad \leq 1 \\ 0 \text{ for } \|x-s\| < 1. \end{cases}$$

Suppose $x_1 = (10,10)$ so $u(x_1) = 1$, and $x_2 = (5,0)$ so $u(x_2) = 0$.

Now $\underset{g}{E}\, u(Y_1) = \underset{g}{E}\, u(Y_2) = 66/132 \cdot 66/50 + 72/132 \cdot 0 = 1/2$ since the area of A is 50 and that of S is 132. Thus,

$$u(x_1) > \underset{g}{E}\, u(Y_i) > u(x_2).$$

Since half the unit circle about x_2 (the support of $g(s|x_2)$) is in A,

$$\underset{\psi}{E}\, u(Y_2) = 1/2 \cdot 66/50 + 1/2 \cdot 0 = .66.$$

Only a slice of area less than 1/2 of the unit circle about x_1 is in A, however; hence

$$\underset{\psi}{E}\, u(Y_1) < 1/2 \cdot 66/50 = .66.$$

Thus $\underset{\psi}{E}\, u(Y_2) > \underset{\psi}{E}\, u(Y_1)$.

The utility function can be rendered differentiable by letting there be a thin zone of width $\varepsilon_1 > 0$ around its boundary in which u(y) slopes (without "corners") from 66/50 down to 0 as y goes from A across this zone into B; by smoothing the corners of A into curves; and by putting x_1 in A (so $u(x_1) = 66/50$) and x_2 just outside the zone in B (so $u(x_2) = 0$ still). If ε is small enough and the curvature at the "corners" sharp enough, the result still goes through in the same manner.

The slope of u within the zone around A can also be slightly perturbed to get slightly but strictly convex sets within each indifference curve, so u would be strictly quasi-concave.

Finally, the uniform sampling distribution could be slightly smoothed out to give a small probability g of being outside the circle $\{x \mid \|x-y\| \leq 1\}$ without changing the result. □

What drives the counterexample is (1) the steepness of the slope of u at x_1 and x_2, along with (2) the placement of x_1 at an almost-corner and x_2 at an almost-flat of the boundary of A. If u were steep enough, the triangular shape could be rounded into an ellipse; if A were eccentric enough, the slope of u could be more gentle.

Counterexample 4: This example is like Counterexample 3, except it uses the somewhat more regular case of elliptical indifference curves, at the cost of added complication. Let S be the rectangle in $\mathbb{R}^2$ bounded by $z_1 = \pm 40$ and $z_2 = \pm 10$. Let A be the interior and boundary of an ellipse given by the formula $z_1^2/1296 + z_2^2/81 = 1$. Let B be the interior of an ellipse given by the formula $z_1^2/1444 + z_2^2/90.25 = 1$. The first ellipse has center (0,0), semi-major axis 72, and semi-minor axis 18, parallel to the z_1 and z_2 axes, respectively. The second is proportional to the first, having axes of lengths 76 and 19. Let the utility function be defined as follows:

$$u(s) = \begin{cases} 100 \text{ if } s \in A \\ 0 \text{ if } s \in S \setminus B \end{cases}$$

and declining linearly from 100 to 98 in $B \setminus A$. Thus for example if $s = (z_1, z_2)$ satisfies $z_1^2/1369 + z_2^2/85.5625 = 1$ then $u(s) = 99$.

Let f be as in Counterexample 3. Let g_1 and g_2 be uniform on S. Suppose $x_1 = (0,9)$ and $x_2 = (-37,0)$ so $u(x_1) = 100$ and $u(x_2) = 99$.

Notice first of all that the unit circle about x_1 lies partly outside B, while that around x_2 lies completely within B, and within that circle utility declines linearly from 100 to 98 as s travels away from the origin. Clearly

$$\underset{\psi}{E}\, u(Y_2) = \int_{\|s-x_2\|\leq 1} u(s) \cdot 1/\pi \, ds = 99.$$

As for $g(s|x_1)$, the boundary in its support dividing the area of zero utility from the area of positive utility is a section of the ellipse which bounds B. It is tangent to the line $z_2 = 9.5$ at $z_1 = 0$. Thus an upper bound on $E\, u(Y_1)$ is provided by multiplying the maximum utility value, 100, by the area of the circle with $z_2 < 9.5$. This leaves three-quarters of the circle's total vertical diameter; a simple integration shows that this area is .8045 of the total area, π. Thus

$$\underset{\psi}{E}\, u(Y_1) < 100 \cdot .8045 = 80.45.$$

So again, $\underset{\psi}{E}\, u(Y_1) < E\, u(Y_2)$. $\sqsubset$

Finally, it is easily seen that the circular contours of the sampling distribution also cannot be dispensed with.

<u>Counterexample 5</u>: In this example, all the conditions of Theorem 6 are met except that the sampling distribution does not have circular contours. For simplicity we let f have a rectangular contour, but it could as well be elliptical or otherwise elongated. Let S be the rectangle in $\mathbb{R}^2$ bounded by $z_1 = \pm 10$ and $z_2 = \pm 10$. Let g_1, g_2 be uniform on S. Let $\alpha = 0$ and let $u(s) = -\|s\|^2$. Let $f(x_i|s) =$ 1/16 on the rectangle bounded by $z_1 = x_{i1} \pm 4$ and $z_2 = x_{i2} \pm 1$, and zero elsewhere. Suppose $x_1 = (0,-2)$ and $x_2 = (-3,0)$ so $u(x_1) = -4$ and $u(x_2) = -9$.

Then $\underset{\psi}{E}\, u(Y_1) = \int_{-4}^{4} \int_{-3}^{-1} -\|s\|^2 \cdot 1/16 \, ds_2 ds_1 = -\,14/3$, while

$$\underset{\psi}{E}\, u(Y_2) = \int_{-7}^{1} \int_{-1}^{1} -\|s\|^2 \cdot 1/16 \, ds_2 ds_1 = -\,1/3$$

so again $\underset{\psi}{E}\, u(Y_1) < \underset{\psi}{E}\, u(Y_2)$. □

REFERENCES

Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee. *Voting*. Chicago: University of Chicago Press, 1954.

Boyd, Richard W. "Popular Control of Public Policy: A Normal Vote Analysis of the 1968 Election." *American Political Science Review* 66 (1972):429-449.

Brody, Richard A. and Benjamin I. Page. "Indifference, Alienation and Rational Decisions." *Public Choice* 15 (Summer 1973):1-17.

DeGroot, Morris H. *Optimal Statistical Decisions*. New York: McGraw-Hill, 1970.

Feller, William. *An Introduction to Probability Theory and Its Applications*, volume 2. 2nd rev. ed. New York: John Wiley and Sons, 1971.

Ferejohn, John A. and Morris P. Fiorina. "The Paradox of Not Voting: A Decision Theoretic Analysis." *American Political Science Review* 68 (1974):525-536.

Fiorina, Morris P. "The Voting Decision: Instrumental and Expressive Aspects." Journal of Politics 38 (1976): 390-413.

Goldberg, Arthur. "Discerning a Causal Pattern Among Data on Voting Behavior." American Political Science Review 60 (1966):913-922.

Hinich, Melvin J. and Peter C. Ordeshook. "Abstentions and Equilibrium in the Electoral Process." Public Choice 7 (Fall 1969):81-106.

Jackson, John E. "Issues, Party Choices, and Presidential Votes." American Journal of Political Science 19 (1975):161-185.

MacQueen, J. and R.G. Miller, Jr. "Optimal Persistence Policies." Operations Research 8 (1960):362-380.

Pomper, Gerald M. "From Confusion to Clarity: Issues and American Voters, 1956-1968." American Political Science Review 66 (1972):415-428.

RePass, David E. "Issue Salience and Party Choice." American Political Science Review 65 (1971):389-400.

Riker, William H. and Peter C. Ordeshook. "A Theory of the Calculus of Voting." American Political Science Review 62 (1968):25-42.

Rothschild, Michael. "Searching for the Lowest Price When the Distribution of Prices is Unknown." Journal of Political Economy 82 (1974):689-711.

Rothschild, Michael and Joseph E. Stiglitz. "Increasing Risk I: A Definition." Journal of Economic Theory 2 (1970): 225-243.

Shapiro, Michael J. "Rational Political Man: A Synthesis of Economic and Social-Psychological Perspectives." American Political Science Review 63 (1969):1106-1119.

Weisberg, Herbert F. "Rational Abstention Due to Satisfaction: A Revised Look at the Effects of Candidate Evaluation on Turnout." Mimeographed. Columbus, Ohio: Department of Political Science, The Ohio State University, 1977.

Chapter 3

DIRECT OBSERVATION OF UTILITY VALUES

In Chapter 2, we modelled the voter's problem as being one of determining the policy positions of the candidates and choosing the preferred one. The utility function was assumed to have a term u(y) which expressed the utility of voting for a candidate at position y; this term could be interpreted either as indicating a direct consumption value of voting for that candidate, or as representing an instrumental value of helping to elect the candidate who promises to deliver y. In either case, it was uncertainty about the true positions y_1 and y_2 which complicated the voter's choice. Imperfect but direct observations of those positions could be made so as to reduce this uncertainty.

Not everyone has imagined voters to be so directly concerned with actual issue positions, though. Even voters who are rational in some sense might use other kinds of information to determine their optimal behavior. Students of politics have written in particular on voters whose preferences between candidates seem to be based on satisfaction with outcomes rather than on the policies expected to produce certain outcomes; the "nature of the times" voter of Campbell et al. [1960] and the "retrospective" voter of Key [1964] and Fiorina [1977, 1979] are classic examples. In this chapter we shall model voters as trying to directly determine the utility level they can expect from each candidate's election, rather than anticipating particular policies. Several other models of voter behavior of this kind will be examined in light of the present model.

1. AN EXTENDED CONCEPTION OF VOTER PREFERENCES

We have previously assumed that voters have preferences over the policy positions available to the candidates. While $u(y)$ could be interpreted as any type of consumption value of voting for the candidate at y, a strict application of the concept of self-interested behavior requires that this $u(y)$ be related to more direct consumption of public or private goods. Interpreted as an instrumental value, $u(y)$ is implicitly derived from the voter's beliefs as to what policies will yield desirable outcomes. In either case, to model a more strictly rational, self-interested voter we should define the fundamental preferences of the individual as preferences over the private and public goods he consumes. Here, public goods should be taken to include not only the usual capital goods and externalities of production and consumption, but also general conditions in the physical, social, and economic environment of the individual. It is due to the possible effect of government policy on these public goods that policies, and hence candidate platforms, take on importance to the self-interested voter.

Formally, if we let z represent the voter's consumption of private goods and w the available bundle of public goods, the voter's fundamental preferences are defined over states of the world (z,w). In order to have induced preferences over candidate positions y, the individual must perceive the connection between y and (z,w). We can construct a simple model of what this connection must look like by letting s represent government policy and λ and u be random variables. Public goods w are determined partly by government policies and partly by random factors: $w = w(s,\lambda)$. Private good consumption may also be affected by public goods, and by government policies such as taxation; so $z = z(w,s)$. Finally, government policies are themselves related to the platform of the winning candidate. Also, however, the complexity of policymaking provides a component which is to the voter at best probabilistic; hence $s = s(y,\mu)$. For y to be related by the voter to his own consumption, then, the determination of w, z, and s must be understood and the distributions of of μ and λ must be estimated.

For the voter who feels he can understand the linkages between platforms, policies, and outcomes and who has subjective probability distributions for λ and μ, u(y) can be derived. Letting u^* be the fundamental utility function defined on states of the world, we have simply

$$\begin{aligned} u(y) &= u^*(z,w) \\ &= u^*[z(w(s(y,\mu),\lambda),s(y,\mu)),w(s(y,\mu),\lambda)]. \end{aligned}$$

The voter has no choice about λ and μ, and his problem is reduced to that of Chapter 2: determine what y_1 and y_2 are likely to be, possibly trading off some of z in the process, and vote for the candidate expected to be best (or abstain from voting). But if the voter has no conception of how z, w, and s are determined or is unable to guess how the random factors might enter that determination, u(y) cannot be defined for a self-interested individual. Thus the sequential search process in which y_1 and y_2 are estimated has no relevance.

2. OBSERVATION OF UTILITY VALUES

If an individual did believe, however, that z, w, and s were determined in some systematic way by y, and that this system did not undergo any drastic changes for likely variations in λ and μ, he might still be interested in what states of the world are likely to be brought about by each candidate. Even without a clear conception of how public goods themselves affect his own consumption (for example, of the relation between z and w), an individual might believe that reliable estimation of the values of $u(y_i)$ is possible. Since the relation is now between i and u rather than between y and u, let us write u_i for the (self-interested) utility that will result from voting for candidate i. Where could information about u_i be obtained? First, the individual may believe that certain actors, perhaps in government or the news media, both understand the determination of w and z and have preferences similar to his own. A pronouncement by such an actor that the election of candidate i will result in a level of satisfaction u_i might be regarded by the voter as being reliable to a significant degree. Conversely, if the actor's preferences were perceived as being extremely different from the voter's own, such information might be equally reliable as a reverse indicator of u_i. The theory of positive and negative reference groups (see the discussion in Campbell, et al. [1960], p. 296) can be interpreted as describing a version of this process. Military officers, former presidents, editorialists, ministers, and opinion leaders of all kinds often perform such a function in political campaigns; and even one or more of the candidates themselves might be seen by the voter as understanding and being honest about policy outcomes and having preferences similar to those of the voter. A second source of information about u_i is the voter's own experience under candidate i's previous terms of office. If things have gone sour during the incumbent president's administration, for example, the voter may choose to vote against him merely on the expectation that he will continue to provide the same kinds of outcomes.

A third, and extremely important, source of information about u_i is available when political parties are a factor in the election for a given office. If the voter believes that candidates from the same party are likely to produce similarly desirable outcomes, he can use his experience under previous officeholders of a given party, or even his memory of the campaigns of previous candidates, to help estimate u_i where i is the current candidate of that party. This belief in the coherence of party outcomes may stem either from an expectation that the policies of future officeholders will be like those of the past for a given party; or just that officeholders of a given party will continue to protect the same interests with the same success as in the past. In the absence of any information about the current candidate himself, the voter's estimate of u_i based on previous information about the party of candidate i can be viewed as the determinant of his "party identification." We shall further explore this idea in the next section.

For some voters at some times, there may not be enough information about the determination of the state of the world for the voter to make use of policy information, so the model of Chapter 2 may be inappropriate. But since direct information on the utility levels to be expected may be available and useful, we may use a model analogous to the previous one. Suppose the individual has a subjective prior density function g_i describing his beliefs about a random variable U_i, the utility which will result if candidate i is elected. By incurring a utility cost c, the individual can obtain a "noisy" observation of the probable utility values from a known sampling density $f(x|u)$, whose parameter is the pair of true utility values which will result if each candidate is elected, $u = (u_1, u_2)$. Again, Bayesian updating is used to incorporate the observation into prior beliefs. Defining

$$v_0(g) = \max \{\underset{g}{E}U_1, \underset{g}{E}U_2\}$$

as the expected value of stopping when the subjective density is g, we can define a functional equation for the voter's sequential search problem exactly analogous to that in Chapter 2:

$$v^*(g) = \max \{v_0(g), \underset{g}{E}v^*(\psi(X,g)) - c\}.$$

If U_1 and U_2 are assumed to lie in a bounded interval of the real line with probability 1, Theorems 1 and 2 on the existence and uniqueness of v^* are exactly as before. Theorems 3 and 4, on the effects of changing search cost, and Theorem 5, on the effects of changing risks, still hold as well. Theorem 6, connecting observations with updated expected values, can now be strengthened somewhat. As before, let R be the reflection mapping about the line $u_1 = u_2$ in $\mathbb{R}^2$, so for any point, $R(u_1,u_2) = (u_2,u_1)$. Whereas in Theorem 6 it was necessary to assume g to be uniform on S x S, we can now simply assume that g satisfies $g(u) > g(R(u))$ whenever $u_1 > u_2$. This is satisfied in particular when g has circular contours on $\mathbb{R}^2$, that is, g_1 and g_2 differ only by the location of their means, both are symmetric, and $\underset{g}{E}U_1 > \underset{g}{E}U_2$. An analogous assumption is made about f.

<u>Theorem 7</u>: Suppose $\underset{g}{E}U_1 > \underset{g}{E}U_2$. For any u such that $u_1 > u_2$, assume

(i) $g(u) > g(R(u))$; and

(ii) $f(s|u) > f(s|R(u))$ for all s such that $s_1 > s_2$.

If the observation x has $x_1 > x_2$, then,

$$\underset{\psi(x,g)}{E} U_1 > \underset{\psi(x,g)}{E} U_2.$$

Proof: Notice that

$$g(u|x) = \frac{g(u)f(x|u)}{\int g(t)f(x|t)dt}$$

and

$$g(R(u)|x) = \frac{g(R(u))f(x|R(u))}{\int g(t)f(x|t)dt} .$$

If $u_1 > u_2$, then, assumption (i) and (ii) imply

$$g(u|x) > g(R(u)|x).$$

Now we have

$$\begin{aligned} \underset{\psi(x,g)}{E}[U_1-U_2] &= \int_{u_1>u_2}(u_1-u_2)g(u|x)du + \int_{u_2>u_1}(u_1-u_2)g(u|x)du \\ &= \int_{u_1>u_2}(u_1-u_2)g(u|x)du + \int_{u_1>u_2}(u_2-u_1)g(R(u)|x)du \\ &= \int_{u_1>u_2}(u_1 - u_2)[g(u|x) - g(R(u)|x)]du \\ &> 0, \end{aligned}$$

the latter since $u_1 > u_2$ and $g(u|x) > g(R(u)|x)$ everywhere in the range of integration.

□

Thus in particular, under much more general assumptions than were necessary for Theorem 6, the voter following an optimal sampling plan will never choose a candidate who was not at least as good as his opponent according to the final observation.

3. RETROSPECTIVE VOTING AND PARTY I.D.

Fiorina [1977] has proposed a general model of the voting decision based, in effect, on observation of utility levels from both the present campaign and past government performance. The latter embodies what Key ([1964], pp. 543-544) termed "retrospective voting," voting so as to punish or reward the incumbent officeholders. While Fiorina's model uses an objective function for the individual voter and the concept of preferences over government programs, it is not an explicitly rational-choice approach in that the objective is not simple maximization of the utility functions, and in that various discounting coefficients which play important and dynamic roles in the decision are completely exogenous. In this section we suggest that the present model of the voter's search problem using direct observation of utility can be used to provide a more strictly rational foundation for the Fiorina model and also to provide a partly endogenous interpretation of the discounting coefficients.

Suppose we let $y^j = (y_1^j, y_2^j)$ represent the positions of the candidates of the two parties during term j for $j = 1,2,\ldots,$ p, p+1, where p is the present and y^{p+1} represents the perceived platforms of the candidates in the current election. Let SS_j denote the status quo in period j, that is, the policies implemented in term j. In Fiorina's model, $SS_j = y_1^j$ or $SS_j = y_2^j$ depending on which party was in office. The voter in Fiorina's model, then, bases his decision on the relative sizes of the following evaluation functions:

$$V_1 = \sum_{j=1}^{p} s_j \alpha_j [u(y_1^{j+1}) - u(SS_j)] \tag{1}$$
$$V_2 = \sum_{j=1}^{p} r_j \alpha_j [u(y_2^{j+1}) - u(SS_j)].$$

Here, s_j, α_j, and r_j are constants. Discounting of the importance of past experiences according to their relevance in the present election is accomplished by $\alpha_j \geq 0$. If party 1 was incumbent in period j, $s_j = 1$; otherwise $0 \leq s_j \leq 1$, so s_j provides an uncertainty discount for the voter's evaluation of what might have transpired _had_ party 1 been in power, based perhaps on party pronouncements during that term as well as on their previous platform. The same role is played by r_j for party 2. Thus, for example, the voter's evaluation of the performance of an incumbent of party 1 enters the evaluation as

$$\alpha_{p-1}[u(y_1^p) - u(SS_{p-1})]$$

and the current platform of party 2, the opposition, appears in the evaluation as

$$r_p\alpha_p[u(y_2^{p+1}) - u(SS_p)]$$

$$= \quad r_p\alpha_p[u(y_2^{p+1}) - u(y_1^p)].$$

The terms $s_p\alpha_p$ and $r_p\alpha_p$ should be given a special interpretation, as discounts for uncertainty about the dependability of the parties' current platforms.

Fiorina's idea of the role of party identification results from breaking down (1) into past and present terms. Let PPE_i stand for the "past political experiences" with party i and let CIC_i represent the "current issue concerns" of the voter with respect to the candidate of party i. Fiorina makes the following definitions:

$$PPE_1 = \sum_{j=1}^{p-1} s_j\alpha_j[u(y_1^{j+1}) - u(SS_j)]$$

$$CIC_1 = s_p\alpha_p[u(y_1^{p+1}) - u(SS_p)]$$

$$PPE_2 = \sum_{j=1}^{p-1} r_j\alpha_j[u(y_2^{j+1}) - u(SS_j)]$$

$$CIC_2 = r_p\alpha_p[u(y_2^{p+1}) - u(SS_p)].$$

Letting γ be a constant representing "an initial bias ... which the individual brings to the political arena (... a direct function of socialization ...)," Fiorina then defines party identification as

$$PID = PPE_1 - PPE_2 + \gamma$$

so that the voter "identifies" with party 1 if PID > 0, with party 2 if PID < 0, and is an independent if PID = 0. In these terms, the voter's decision is to choose candidate 1 if $V_1 > V_2$, that is, if

$$PID + CIC_1 - CIC_2 \geq \gamma.$$

Fiorina points out (Fiorina [1977], p. 611) that, although it uses several of the elements of a rational choice theory of voting (the utility function u defined over issue positions and the basing of voter choice on his preferences and on the real life actions of political actors), his model is not stated as a rational choice theory. The criterion of the voter is based on a comparison of V_1 and V_2 and not on a direct maximization of the given utility function. Also, the voter makes use of his information by simply adding the results of his past observation using some discount factors exogenous to the model; such an approach may be characterized as a rule of thumb, as opposed to an explicit rational analysis of the use of information. But are there any features in the model which cannot be reinterpreted as being rational using the terms Fiorina has defined? Notice first of all that the positions and policies y^j never appear in the model except as arguments of u. As can be seen by comparing Theorem 6 and the ensuing counterexamples of Chapter 2 with Theorem 7, the observation of positions themselves entails considerations of the riskiness of those positions and the consistency of observations which do not arise when utility values alone are observed. Those considerations do not appear in the Fiorina model; in effect, Fiorina's voter is concerned with satisfaction only and not with how the candidates propose to provide it. Such a conception is in keeping with the spirit, as well as the letter, of the Fiorina model. This is particularly evident from Fiorina's comparison of his model with that of Key [1964]:

The bracketed term weighted by α_{p-1} is a bias, a symbolic pat on the back or kick in the pants from an electorate that is a "rational God of vengeance and reward" (Key [1964], p. 568). If the citizen has prospered under the incumbent, he enters the voting booth predisposed toward the incumbent, ceteris paribus. If the citizen has suffered, the challenger might capture his vote even with an inferior campaign platform (Fiorina [1977], p. 604).

Second, Fiorina explicitly treats the terms $[u(y_i^{j+1}) - u(SS_j)]$ as uncertain or noisy information about the candidate's future performance. While the performance of past incumbents is formally just a "bias" in the citizen's voting decision, Fiorina also refers to it as the "hardest bit of information [the voter] has" (Fiorina [1977], p. 603) about future changes in his welfare, and refers also to Downs' [1957] argument along the same lines. However, even the performance of the present incumbent is not a precise indicator of future well-being, so α_p does not carry all of the weight of the α_j coefficients.

In short, Fiorina's $u(y_i^{j+1})$ terms are noisy observations of future utility levels, exactly like the observations x_i in the voter's sequential search problem with direct observation of utility levels, where party cohesion makes information about past candidates of a given party serve as information about the latest candidate as well. The analogy is even clearer if we think of the x_i as observations of the <u>change</u> in utility which may be wrought by candidate i, a modification which does not alter the model of voter search. Letting x_i^j represent the observation of party i from period j, we can rewrite Fiorina's evaluation functions as

$$V_1 = \sum_{j=1}^{p} s_j \alpha_j x_1^j$$

and similarly for V_2. Fiorina makes "an obvious technical point: the simple additive structure (of V_i) is only one among many formal structures which could encompass" his concepts of the voting decision (Fiorina [1977], p. 611). We might also, then, write $V_i = \underset{\psi}{E} U_i$ where ψ is the posterior distribution of the utility change U, updated by observations $x^1, \ldots, x^p$. The fact that the observations originate in different time periods under different incumbency conditions means that the sampling distribution of x_1^j may be different from that of x_2^k. The Bayesian updating process uses those observations differently; intuitively, it puts more weight on observations of higher precision, providing an obvious analogy to the discounting coefficents s_j, r_j, and α_j. Thus the updating process alone provides a close rational choice analogy to

Fiorina's model, exactly in the spirit of Zechman's [1978] model of party identification as an informational cue updated by observation of policy positions and platforms. Both approaches, however, make the simplifying assumption that exactly one observation is obtained at each period, leaving unexplained how that information is chosen over alternative information and how the precision of the observations is estimated. We can apply the fuller model of the voter's sequential search problem to gain additional understanding of these points, and of the simultaneous use of party identification and current issue concerns.

Let us consider the voter at a single election. His general search problem is to either: immediately choose a candidate using his prior; sample from period $j < p$, where the search cost c_j represents merely a cost of evaluating his experience in light of his current situation; or sample from the current campaign at cost c, as before. We assume, of course, that the voter holds a subjective sampling probability distribution $f^j(x|u)$ for each period j. As pointed out in the final section of Chapter 2, this problem has many of the features of the general model, but the model as given tells us nothing about the question of choosing from among several sampling distributions. We can go further with the application of the voter search model by making two simplifying assumptions in the spirit of Fiorina's model: (1) for each $j < p$, exactly one observation X^j is available; and (2) the costs c_j for $j < p$ are essentially zero. Thinking of the voter's "original" prior g^0 as being analogous to Fiorina's γ term representing pre-political information or learning through socialization, we then have the voter automatically considering $x^1,\ldots,x^{p-1}$ as observations and forming his "prior" for this election $g = \psi[(x^1,\ldots,x^{p-1}),g^0]$. If both candidates are new, this prior represents the voter's party identification: he identifies with the party i for which $EU_{g\,i}$ is greatest.

The voter's remaining problem, then, is to sample from the current campaign and update g, exactly as in the general model. The fact that several observations are taken in the current period should pose no conceptual problem: consider the case where $g(u)$ and $f(x|u)$ are both normal distributions, having means and precisions (μ,ρ) and (u,τ) respectively. If n observations are taken, the posterior distribution has precision $\rho + \tau n$ and mean

$$\frac{\rho\mu + \tau x^1 + \ldots + \tau x^n}{\rho + \tau + \tau + \ldots + \tau} = \frac{\rho\mu + \tau \sum_{k=1}^{n} x^k}{\rho + n\tau}$$

$$= \frac{\rho\mu + n\tau\bar{x}}{\rho + n\tau}$$

where $\bar{x}$ is the mean of the sample (see DeGroot [1970], p. 167). This is exactly as if one observation, $X^p = \bar{x}$, had been taken from a distribution with precision $n\tau$. Thus the analogy of Fiorina's α_p term is determined by the sequential sampling behavior of the voter: ceteris paribus, each additional observation increases the precision of the "single observation" $\bar{x}$ and thus its weight in determining the posterior mean. This interpretation could of course be extended to multiple observations from previous periods, as well.

We can now look at several empirical types of voting behavior examined by Fiorina in terms of the voter's sequential sampling behavior. The "issue voter" is one for whom c is low enough and (in the case of normal f and g) τ high enough, relative to the prior g, that the voting decision depends mainly on observations from the current campaign. Empirically, we would expect such voters to exhibit one or more of the following characteristics: a high level of attention to the campaign, a large amount of information about the candidates, a strong opinion about one or the other candidate's performance in some policy area, and a strong preference between possible states of the world perceived to be at issue in the current campaign. The "irresponsible voter" may exhibit the opposite relations between parameter values (high cost, low precision or interpretability of signals) and the opposite characteristics of those listed above; or he may simply have a very precise estimate of U_i from past political experience, causing current issue concerns to pale by comparison. Either of these irresponsible voters could be a party-line voter, but the former could also include voters with weak or nonexistent party identification who simply derive a high consumption value from voting for the preferred candidate, however vague the preference. Another kind of party voter, typified by Fiorina's hypothetical "I like Reagan, but Hoover took Dad's farm" (Fiorina [1977], p. 613) is characterized as basing his choice on perhaps a single previous observation even in the face of reasonably coherent information about the present candidates. Such a case could result from some X_i^j being extremely bad (or good) but having reasonably high precision, so that g is so skewed toward one party that even a moderate amount of current information cannot swing it back. The Michigan school's "nature of the times" voter (Campbell et al. [1960]) is among the "irresponsible" voters described here, as well.

Fiorina's explanation ([1977], pp. 620-622) of incumbent advantage has a drawback which can be explained clearly using the present model. When candidate 1 is an incumbent, typically $s_j > r_j$; hence a given level of utility improvement by the incumbent will outweigh a larger level potentially available from the opposition due to uncertainty discounting. The longer a candidate (or party) has been in power and providing such mediocre benefits, the larger his advantage in the calculations of V_1 and V_2. This seems to indicate that all voters are risk-averse, preferring a lower but less risky expected payoff, in rational choice terms. But no such assumption was made by Fiorina either about u or in the construction of the V_i. Under Bayesian updating, starting say from a prior with $g_1 = g_2$, updating using the incumbent's high-accuracy signals may yield a more precise posterior than for the challenger whose signals are from a riskier distribution. Under the model of direct observation of utility levels, however, the voter is by definition risk-neutral, since utility is a linear function on the state space (just utility values). Analogously, Fiorina's V_i is just a linear function of the X_i^j observed. The present formulation, however, does not support such an uncertainty-based interpretation of incumbency advantage; that interpretation results from Fiorina's additive and multiplicative structure of V_i which, as we noted earlier, he does not view as a salient feature of the model.

For the voter in a single campaign, we have rendered Fiorina's "psychological" parameters s_j, r_j, α_j, and γ as parameters of the search model. In discussing behavior over several elections, other exogenous changes are suggested by Fiorina, some of which cannot be escaped using the present model. For example, if a voter should decide in election p'(that is, the election after term p') that the events of term j are more highly relevant than they were at election p < p', we must posit a change in the corresponding precision of X^j relative to the precision of other observations just as Fiorina must resort to a relative change in α_j'. And of course Fiorina's voter must always produce his new values of α_p and r_p as each new election arises, just as ours must come up with $f(x^p|u)$ and c. Suppose we assume in general that at each election p, the voter of Fiorina's model acquires a value of α_{p-1} which does not necessarily resemble the α_{p-1} value he had at election p-1. In terms of the voter search model this is completely reasonable since evaluating past information at election p is likely to be very different from observing present campaign information in election p-1. Such a dynamic assumption is necessary in order to account for voting against one's party identification without changing that identification. Another source of such behavior should arise when a new candidate is being evaluated on the basis of past information about other candidates of his party; the past information may be viewed as irrelevant in evaluating an unusual candidate, just as information about a deviating candidate will be of little use in the future to evaluate a more conventional candidate. Again, Fiorina's model consigns all such action to changing of the α_j values over time, just as the current model consigns it to the formation of new sampling distributions. Other dynamic phenomena are subsumed more effectively by the two models, such as: the increasing strength of party identification with length of party experience, due to successive updating of g to more and more precise posteriors; changing party identification, due to an observation which changes the sign of $\underset{g}{E}U_1 - \underset{g}{E}U_2$; and "critical elections," in which some objective conditions of low-cost information and highly salient issues cause many voters to change party identification. In principle, though, the voter search model is more complete than Fiorina's model in accounting for the dynamic parameter changes mentioned above. The reliability of sampling distributions is partly just another set of information about the political environment. The consistent (one might even say relentless) rational-choice and information-economizing character of the present model means that it could incorporate information about sampling distributions just as it already employs information about utility levels.

Such concepts begin to stray from the realm of measurable characteristics and testable hypotheses, however, at least for the present. The first empirical question one might ask when confronted with the voter search model is whether voters actually behave as information-economizing expected utility maximizers would behave. How might one test this model? Conveniently, one answer lies in the testing of the Fiorina model. Suppose that our voter, while not sufficiently cognizant of issue positions and policies to use those in evaluating candidates, at least has some notion about how his utility is affected by different components of the public goods bundle which helps make up the state of the world. When asked whether the government has performed well in the area of unemployment or foreign relations, such a voter could likely give a coherent answer based on how recent changes in unemployment and foreign affairs have affected his own feeling of well-being. Such answers in political surveys thus constitute observations on the x^p values of individuals, and we can test whether these values relate to vote decisions and party identification in the appropriate ways. This process describes the testing of the present model or of the Fiorina model. Fiorina [1979] has in fact carried out such tests relating "retrospective evaluations" to voting and the development of party identification.

Other factors which would be relevant in testing the present model would include the voter's levels of political information and interest, which do not appear explicitly in Fiorina's model but are more central to the voter search model which, we claim, underlies it. In effect, the Fiorina model of party identification and voting is the tool for setting up empirical tests of the rational choice model of voter search.

Having attached the voter search model to Fiorina's model, we can apply Theorems 1 through 7 to obtain further results about the logical consistency and behavior of Fiorina's hypothetical voters. For the latter model, Theorems 1 and 2 now imply that such information-seeking and optimizing behavior by voters is a consistent and well-defined way to make fully rational voting decisions. Theorems 3 and 4 show how such voters respond to different costs of information, a problem not explicit in Fiorina's model but implicit in the fact that his voters are limited in their use of information (indicating that information is costly). Exactly as in the voter search model, Theorem 5 relates the risk faced by the voter to his sampling behavior, and Theorems 6 and 7 relate observations to choices. For a rational voter who does not make use of information from the current campaign, Theorem 7 in particular shows that under certain conditions his evaluation of the present incumbent is of primary importance in predicting his voting choice. The present model and Fiorina's [1977, 1979] work can thus be viewed as mutually reinforcing: the voter search model shows the Fiorina model to be really based on a strictly rational-choice framework, and to be internally consistent as a theory of rational voter behavior; while Fiorina's work clearly shows the connection between the voter search model and various empirical features of voting behavior, especially party identification and retrospective voting.

4. OTHER VERSIONS OF DIRECT OBSERVATION OF UTILITY LEVELS

The Bayesian voting model of Powell [1974], discussed in Chapter 1, is a precursor of the present model in that the voter is assumed to observe, at a cost, possible utility levels from the candidates before making a decision. There, the voting decision itself is based on the Riker-Ordeshook [1968] model: if P represents the probability of affecting the outcome of the election, c_v the cost of voting, D the "citizen duty" term, and u_1 and u_2 the <u>true</u> utility values of the candidates, the goal is to vote for candidate 1 if $u_1 > u_2$ and $P \cdot (u_1 - u_2) - c_v + D > 0$; to abstain if $P \cdot |u_1 - u_2| - c_v + D \leq 0$; and to vote for candidate 2 if $u_2 > u_1$ and $P \cdot (u_2 - u_1) - c_v + D > 0$. This is essentially identical to the present model using abstention due to indifference (see Chapter 2, Section 7) in which the P and D terms play no role, since they do not really matter in the Powell model either. Allowing the voter to sample individually from each of the two candidates' signals, Powell attempts to show the effects on sampling and voting of variations in sample variance and the information-cost function, and to characterize the optimal number of observations of each candidate. However, Powell does not take a sequential sampling and dynamic programming approach as in the present study; her results are based mainly on an optimal sample-size approach and necessitate the addition of rules-of-thumb for the voter to estimate the true values u_1 and u_2 even after applying Bayesian updating. (For a fuller discussion of Powell's model, see Chapter 1.) Conceptually, though, her model of the voter's problem is identical to that in the present study where direct observation of utility is assumed.

Plott and Wilde [1979] have presented a model which makes several simplifying assumptions beyond those used in the present model, and obtains considerably stronger results. Their model is an attempt to study the behavior of a consumer who must depend upon sellers for advice about which of two products would actually be preferable. Suppose that a and b are two products whose prices are fixed. In "demand state" A, the consumer will receive utility

$v(a|A)$ if he purchases product a, and $v(b|A)$ for product b, where $v(a|A) > v(b|A)$. Likewise for demand state B, $v(a|B) < v(b|B)$. The consumer is assumed to have a subjective prior probability q that he is really in state A and (1-q) that he is in state B. For a cost c, he can ask a seller to advise him what the true demand state probabilities are; let $f(p|y)$ represent the consumer's belief of the probability a seller will advise him that (p, 1-p) are the probabilities when the true demand state is $y \in \{A,B\}$. (A simplification which does not alter the results is to let the seller advise only $p = 1$ or $p = 0$, that is, to say what he thinks the demand state is.) Using a few regularity assumptions on f, Plott and Wilde derive the following properties of consumer behavior when the consumer engages in optimal sequential sampling using Bayes' rule to update q:

(1) the consumer's dynamic programming functional equation has a unique continuous solution;

(2) under the optimal sampling plan, the consumer stops and chooses a strictly over b if and only if his last observation has $f(p|A) > f(p|B)$, and is indifferent if and only if the last p has $f(p|A) = f(p|B)$;

(3) under an optimal sampling plan in which at least two observations are taken, the consumer buys a if and only

if the last two predictions p_1 and p_2 "favor" a in that $f(p_1|A) \cdot f(p_2|A) > f(p_1|B) \cdot f(p_2|B)$, and is indifferent if and only if equality holds.

These last two results are in exactly the same vein as Theorems 6 and 7 in the present study: they relate choices to final observations. We can put the Plott-Wilde model into an electoral context by letting products a and b be candidates 1 and 2 and letting the demand states be $u_1 > u_2$ and $u_1 < u_2$. Here, then, the voter does not even concern himself with utility levels of outcomes, but solely with which candidate is better; the utility from voting for the better candidate is not dependent on <u>how much</u> better he is. The voter's utility-level observations become simply information on which candidate is likely to be the better one. Theorem 7 can then be restated: under an optimal sampling plan, the voter stops observing and chooses candidate 1 if and only if his last observation tells him that candidate 1 is more likely to be the better candidate. Furthermore, result (3) above draws a new connection between the last <u>pair</u> of observations and the final choice. Although the highly simplified structure of the Plott-Wilde model essentially makes the same assumptions as were required in Theorem 7, the substantive significance of this result becomes clearer; in an optimal sampling process, the last observations in effect determine the final choice of candidates.

As Plott and Wilde ([1979], p. 6-7) point out, the fact that both products and information are supplied by sellers of the products suggests that, in a complete model of such a market, the subjective sampling density $f(p|y)$ should be endogenously determined. This concept is highly relevant to modelling of the electoral process as well. The candidates, inasmuch as they influence the information acquired by voters, will of course wish to manipulate that information to their own ends. However, a complete loss of voter confidence in campaign information means a complete loss of control of that information for the candidates, which will likely not be an optimal state of affairs for either side of the electoral "market." Although it has not been possible so far to directly develop the voter search model into a more two-sided model, we shall in the next chapter examine indirectly some of the effects of uncertainty (on the part of both voters and candidates) on candidate strategies.

REFERENCES

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. The American Voter. New York: John Wiley and Sons, 1960.

DeGroot, Morris H. Optimal Statistical Decisions. New York: McGraw-Hill, 1970.

Downs, Anthony. An Economic Theory of Democracy. New York: Harper and Brothers, 1957.

Fiorina, Morris P. "An Outline for a Model of Party Choice." American Journal of Political Science 21 (1977):601-626.

__________. Retrospective Voting in American National Elections. Unpublished manuscript. Pasadena, Calif.: Division of Humanities and Social Sciences, California Institute of Technology, May 1979.

Key, V.O. Jr. Politics, Parties, and Pressure Groups. 5th rev. ed. New York: Thomas Y. Crowell, 1964.

Plott, Charles R. and Louis L. Wilde. Untitled manuscript. Pasadena, Calif.: Division of Humanities and Social Sciences, California Institute of Technology, June 1979.

Powell, Lynda Watts. "A Bayesian Model of Individual Voting Decisions in Two-Candidate Elections." Mimeographed. Albany, N.Y.: Graduate School of Public Affairs, State University of New York at Albany, 1974. (Forthcoming in Journal of Mathematical Sociology 6 (1978).)

Riker, William H. and Peter C. Ordeshook. "A Theory of the Calculus of Voting." American Political Science Review 62 (1968):25-42.

Zechman, Martin J. "Dynamic Models of Voting Behavior and Spatial Models of Party Competition." Institute for Research in Social Science Working Papers in Methodology, Number 10. Chapel Hill, N.C.: University of North Carolina, 1978.

Chapter 4

UNCERTAINTY, CANDIDATE GOALS, AND MEDIAN VOTER OUTCOMES IN SPATIAL MODELS OF ELECTORAL PROCESSES

In most spatial models of electoral processes, candidates are assumed to pursue office with a single-minded fixation on winning office or maximizing the power of office through achieving a large mandate. Voters, for their part, are usually accorded accurate enough information to vote for the candidate whose platform is actually best for them, provided they do not abstain. One common result in such models is that, aside from instances where multiple optima occur or where criterion functions are not symmetric between the candidates, Nash equilibrium candidate strategies require at least the winning candidate to adopt a median position when equilibrium exists. Even when policy-related candidate goals have been studied [Wittman, 1977] or where candidates face an uncertain election outcome, this median result is still a regular feature of candidate equilibrium. Usually, in fact, an equilibrium pair of strategies requires <u>both</u> candidates to adopt such a strategy.

In this paper we will examine these situations of policy-oriented candidates and candidates with imperfect information about election outcomes in more detail. In addition, we will draw some simple conclusions about candidates having both the policy outcome and the election outcome in their preference domains. Of the four basic combinations of policy or office orientation with certainty or uncertainty about outcomes, three lead to median voter results; in the case of policy-oriented candidates with uncertainty, however, we will see that no pair of strategies in which two candidates take identical platforms can be an equilibrium strategy pair. This result can be immediately extended to a mixed-motive situation as well, so that if policy preferences and uncertainty are believed to play any role in an election, the tendency of candidates not to converge to the median must be present.

1. OFFICE-ORIENTED CANDIDATES WITH PERFECT INFORMATION

The equilibrium nature of convergent candidate strategies and the importance of the median position in a symmetric, two-candidate election was first noted by Hotelling [1929]. Expanding on the spatial modeling approach, Downs [1957] argued that convergent strategies would dominate if the (unidimensional) distribution of opinion were concentrated around a single point, for example if the distribution were unimodal; but in the case of, say, a bimodal distribution, two candidates interested only in winning elections could be expected to locate at the two modes. Plott [1967] demonstrated that in an issue space of more than one dimension, the existence of any equilibrium position is problematic, a point which has been elaborated upon by Schofield [1978]. For several types of voter distributions under which majority-rule equilibria may exist, however, Davis, Hinich, and Ordeshook [1970] showed that convergence is more likely than Downs had thought. Even with a bimodal distribution of citizens' preferred points, plurality-maximizing candidates will still converge to the mean of the distribution.

For the same type of spatial model as is discussed in the Davis, Hinich, and Ordeshook piece [1970] and previous Davis and Hinich papers [1966, 1967, 1968], conditions under which a nonconvergent candidate equilibrium may occur are given in Hinich and Ordeshook [1970]. These include (1) vote maximization, rather than plurality maximization, as the candidates' goal, along with (2) abstention by citizens according to how small the utility of their most preferred candidate is, called "abstention due to alienation." Under these two assumptions, provided that individual voters are sufficiently sensitive to alienation and that the tails of the distribution of citizens' preferred points are thick enough, candidates moving toward the mean of a symmetric distribution may at some point lose more votes through absention than they gain through being close to more voters. Note that even a unimodal distribution of voters may satisfy these conditions.

Candidates may of course fail to converge when no majority-rule equilibrium exists. Kramer [1975, 1977] examines a dynamic model of candidate strategies in this case and finds that, over time, platforms of candidates following "vote-maximizing trajectories" converge to a relatively small minmax set. However, McKelvey [1976, 1977], Cohen [1977], and Cohen and Matthews [1977] show that in general the "top cycle" of platform positions is larger than the minmax set and may in fact lead the candidates anywhere in the issue space even if they have reached the minmax set.

Clearly, if there are multiple optima in a candidate's objective function, as might happen for instance if abstention exactly balanced vote-switching along some path in the issue space, equilibrium might exist even though the candidates need not converge. This situation might arise under mixed plurality- and vote-maximizing strategies on the part of a candidate or under certain combinations of abstention functions and voter distributions. Finally, even under perfect information there may be constraints on candidates rendering them unable to approach an existing convergent equilibrium. These may take the form of institutional constraints, as when the candidates must gain nomination or resources from party activists; or they may spring from the presence of issues which are not pure policy-choice issues, as for example when the very identity of a candidate has an effect on foreign affairs which is beyond that candidate's control. (On these points, see also the discussion of Davis, Hinich, and Ordeshook [1970], pp. 429-430, 432, 442; and the discussion of valence issues in Stokes [1963]).

Despite these cases in which convergence may not occur, one lesson of the spatial modeling literature in which information is complete and candidates seek only electoral success is that, in the abstract, a democratic choice process carries with it pressures toward convergence. Particular aggregate abstention properties or situations with nonunique optima do not detract from this general conclusion. The generic absence of equilibrium is mitigating, but here as well we have the dynamic but temporary convergence properties of Kramer [1975]. In addition, the one-dimensional model is sometimes directly applicable; and the presence of institutional constraints such as committee jurisdictions can bring about a multidimensional equilibrium which corresponds to the case of one dimension (see Shepsle [1977, 1978]). Thus it still is significant to know whether changing the traditional assumptions about candidate goals or the information environment will change the conclusions about platform equilibria found in the theoretical literature on democratic choice.

2. POLICY-ORIENTED CANDIDATES UNDER PERFECT INFORMATION

Perhaps more surprising, on the surface, than the convergence of candidates when opinion is distributed bimodally is the apparent convergence tendency when candidates are interested primarily in the policy which is eventually implemented. Let S be a convex issue space, $S \subseteq \mathbb{R}^n$; let f(x) describe the distribution of voters' preferred points; and let candidates 1 and 2 have preferred points y^1 and y^2 respectively. We consider first the simple case where n = 1. Suppose that the utility function u^i of a citizen i is restricted only to be a decreasing function of distance from his preferred point x(i). But, let us depart from the traditional spatial model by supposing that each candidate has a similar utility function $v^1(\theta)$, $v^2(\theta)$ where θ is the policy platform of the winning candidate, which is eventually implemented. Let m be the median of f. If $y^1 < m < y^2$, assuming that all citizens vote, then the only Nash equilibrium is where one candidate takes the position $\theta^j = m$. For suppose $\theta^2 < m$; then candidate 2 will not be content to be anywhere such that $\theta^2 \geqslant m + (m - \theta^1)$ since it is possible otherwise to win with a position which he prefers to θ^1. But if $\theta^2 > m$, the same argument applies to candidate 1. Finally notice that $\theta^1 > m$ or $\theta^2 < m$ is not satisfactory to the winning (or tying) candidate, since by adopting m he could either lose to a preferred platform or still win with a better platform. Only when $\theta^2 = m$ or $\theta^1 = m$ does neither candidate have any incentive to move given that the other will not. In the case where both y^1 and y^2 are on the same side of m, of course, this does not hold; but in a setting where candidates are fighting it out over what policy is implemented, the latter does not seem the most important case. In a non-dynamic setting, then, we would predict <u>both</u> candidates to adopt m.

It is apparent that many aspects of traditional spatial models apply to the case of policy-oriented candidates with perfect information. With $y^1 < m < y^2$, clearly Nash equilibrium will occur when neither candidate can improve a loss to a tie or win, or a tie to a win, by moving closer to the other. In particular, the model of <u>vote-</u>maximizing candidates in the presence of abstention due to alienation does <u>not</u> apply. Policy-oriented candidates, here, are interested indirectly in winning (and hence in plurality) since it determines what policy is implemented; they are in no sense interested in maximizing votes in the face of abstention. Thus in the policy-oriented case there is no nonconvergence result corresponding to that of the earlier abstention-due-to alienation assumption.

The extension to arbitrary dimension n is fairly straightforward, with difficulties arising from (1) existence of a majority-rule equilibrium, which we will now term a <u>total median</u> μ of f (after Hoyer and Mayer [1974]); and (2) a more complicated requirement of policy disagreement between the candidates, generalizing the requirement $y^1 < m < y^2$ of the one-dimensional case. We will assume throughout that candidates and voters have type I preferences (circular indifference curves) over S. Define the <u>plurality function</u> $\phi(\theta^1, \theta^2)$ to be the proportion of citizens voting for candidate 1 minus the proportion voting for candidate 2. Given a position $\bar{\theta}^1$ for 1, the shape of f(x) determines the function $\phi(\bar{\theta}^1, \cdot)$ and in particular the contour curve of $\phi(\bar{\theta}^1, \cdot)$ which passes through $\bar{\theta}^1$. This curve bounds the set of positions θ^2 such that $\phi(\bar{\theta}^1, \theta^2) < 0$, i.e., candidate 2 wins. Let C^2 be the straight line segment between μ and y^2. This set represents points at which θ^2 would tie θ^1 while maximizing $v^2(\theta^2)$, given θ^1 at various distances from μ. C^1 is defined similarly.

Suppose $\phi(\bar{\theta}^1, \bar{\theta}^2) > 0$. $(\bar{\theta}^1, \bar{\theta}^2)$ is a Nash equilibrium strategy pair if and only if: (1) there is no θ^1 such that $\phi(\theta^1, \bar{\theta}^2) > 0$ and $v^1(\theta^1) > v^1(\bar{\theta}^1)$; and (2) there is no θ^2 such that $\phi(\bar{\theta}^1, \theta^2) \leq 0$ and $v^2(\theta^2) > v^2(\bar{\theta}^1)$. A similar definition holds for $\phi(\bar{\theta}^1, \bar{\theta}^2) < 0$.

If $(\bar{\theta}^1, \bar{\theta}^2) = 0$, Nash equilibrium has been reached if and only if: (1) $v^1(\theta^1) > v^1(\bar{\theta}^1) \Rightarrow \phi(\theta^1, \bar{\theta}^2) < 0$, and similarly for candidate 2; and (2) $\phi(\theta^1, \bar{\theta}^2) > 0 \Rightarrow v^1(\theta^1) \leq t^1(\bar{\theta}^1, \bar{\theta}^2)$, where t^1 is the utility of a tie election for candidate 1, and similarly for candidate 2 (and t^2). (Note that if ties are decided by a coin toss and candidate j is risk-neutral, $t^j(\bar{\theta}^1, \bar{\theta}^2) = (1/2)[v^j(\bar{\theta}^1) + v^j(\bar{\theta}^2)]$.) In the case $\bar{\theta}^1 = \bar{\theta}^2$, of course, we take $t^j = v^j(\bar{\theta}^1) = v^j(\bar{\theta}^2)$ for $j = 1,2$ since winning has no intrinsic value.

Now suppose f is a symmetric distribution. Its mean μ is a total median, and it is easy to see that if $\bar{\theta}^1 = \bar{\theta}^2 = \mu$ then $\bar{\theta}^1, \bar{\theta}^2$ is a Nash equilibrium: moving away from μ means losing instead of tying. Provided that a condition of sufficient disagreement between the candidates relative to f is met, $\bar{\theta}^1 = \mu$ and/or $\bar{\theta}^2 = \mu$ are the only Nash equilibria. The condition is just that μ, y^1, and y^2 do not lie on a single line unless μ is between y^1 and y^2; that is, the candidates' utility gradients at μ do not have the same direction. In the present case we can write this condition as $\frac{y^1 - \mu}{\|y^1 - \mu\|} \neq \frac{y^2 - \mu}{\|y^2 - \mu\|}$

<u>Lemma</u>: If f is symmetric then $\phi(\theta^1,\theta^2) = 0 \Longleftrightarrow \|\theta^1 - \mu\| = \|\theta^2 - \mu\|$.

<u>Proof</u>: $\phi(\theta^1, \theta^2) = 0 \Longleftrightarrow$ there is some vector h such that

(1) $$\int_{\{x|x\cdot h \geq 0\}} f(x)\, dx = \int_{\{x|x\cdot h \leq 0\}} f(x) dx$$

since we assumed type I preferences. This h is normal to a hyperplane H satisfying

(2) $$(1/2)(\theta^1 + \theta^2) \in H; \text{ and}$$

(3) $$(\theta^1 - \theta^2) \cdot (x-y) = 0 \quad \forall\, x,y \in H,$$

That is, H is the perpendicular bisector of the line segment from θ^1 to θ^2. Because h satisfies (1), H is by definition a partial median. Since μ is a total median, $\mu \in H$. By (2), $[\frac{1}{2}(\theta^1 + \theta^2) - \mu] \varepsilon H$. By (3), $\frac{1}{2}(\theta^1 - \theta^2) \cdot [\frac{1}{2}(\theta^1 + \theta^2) - \mu] = 0$, that is, $[\theta^1 - \frac{1}{2}(\theta^1 + \theta^2)] \cdot [\frac{1}{2}(\theta^1 + \theta^2) - \mu] = 0$. We can therefore apply Pythagoras' law:

$$||\theta^1 - \mu||^2 = ||\theta^1 - \tfrac{1}{2}(\theta^1 + \theta^2)||^2 + ||\tfrac{1}{2}(\theta^1 + \theta^2) - \mu||^2, \text{ and likewise}$$

$$||\theta^1 - \mu||^2 = ||\theta^1 - \tfrac{1}{2}(\theta^1 + \theta^2)||^2 + ||\tfrac{1}{2}(\theta^1 + \theta^2) - \mu||^2.$$

But $\theta^1 - \frac{1}{2}(\theta^1 + \theta^2) = \frac{1}{2}(\theta^1 - \theta^2) = \frac{1}{2}(\theta^1 + \theta^2) - \theta^2$; their norms are equal, so

$$||\theta^1 - \mu||^2 = ||\theta^2 - \mu||^2 .$$

□

<u>Theorem 8</u>: Suppose f is symmetric, and $\frac{y^1 - \mu}{||y^1 - \mu||} \neq \frac{y^2 - \mu}{||y^2 - \mu||}$.

Then $(\bar{\theta}^1, \bar{\theta}^2)$ is a Nash equilibrium if and only if $\bar{\theta}^1 = \mu$ or $\bar{\theta}^2 = \mu$.

<u>Proof</u>: Suppose $\bar{\theta}^1 \neq \mu$ and $\bar{\theta}^2 \neq \mu$, and consider the case where $\phi(\bar{\theta}^1, \bar{\theta}^2) > 0$. By moving to the point θ^2 such that $||\bar{\theta}^1 - \mu|| = ||\theta^2 - \mu||$ and such that distance from y^2 is minimized (so he is on C^2), candidate 2 could achieve a tie (see the Lemma). He will prefer to do so, i.e. $v^2(\theta^2) > v^2(\bar{\theta}^1)$, (and hence $(\bar{\theta}^1, \bar{\theta}^2)$ will not be an equilibrium), unless that point <u>is</u> $\bar{\theta}^1$. But in that case, candidate 1 could maintain his winning position but gain utility by moving to θ^1 such that $||\theta^1 - \mu|| = ||\bar{\theta}^1 - \mu||$

and $\theta^1 = \alpha\mu + (1 - \alpha)y^1$ (i.e. $\theta^1 \in C^1$), that is, the closest point to y^1 which does exactly as well as $\bar{\theta}^1$. By assumption, this θ^1 will not be the same as θ^2. Hence $(\bar{\theta}^1, \bar{\theta}^2)$ is not an equilibrium. A similar proof holds if $\phi(\bar{\theta}^1, \bar{\theta}^2) < 0$.

Next suppose $\phi(\bar{\theta}^1, \bar{\theta}^2) = 0$. By Lemma 1, $\|\bar{\theta}^1 - \mu\| = \|\bar{\theta}^2 - \mu\|$. By the latter part of the previous argument, both $\bar{\theta}^j$ lie in C^j, that is, on a straight line from y^j to μ. But there is some ε such that if $\theta^2 = \varepsilon\mu + (1 - \varepsilon)\bar{\theta}^2$ then $\phi(\bar{\theta}^1, \theta^2) < 0$ while $v^2(\theta^2) > v^2(\bar{\theta}^1)$, i.e.,

$$\|y^2 - \varepsilon\mu - (1 - \varepsilon)\bar{\theta}^2\| < \|y^2 - \bar{\theta}^1\|.$$

To find such an ε, just let α be the number such that $\|(\alpha\mu + (1 - \alpha)\bar{\theta}^2) - y^2\| = \|\bar{\theta}^1 - y^2\|$. We know that $\alpha > 0$ because $\|\bar{\theta}^2 - y^2\| < \|\bar{\theta}^1 - y^1\|$. If $\alpha > 1$, take any $0 < \varepsilon < 1$; if $\alpha < 1$ take $\varepsilon = \frac{\alpha}{2}$.

Thus $(\bar{\theta}^1, \bar{\theta}^2)$ is never a Nash equilibrium when neither is equal to μ.

Conversely, suppose $\bar{\theta}^1 = \mu$. Because μ is a total median, the best candidate 2 could do would be to tie $\bar{\theta}^1$ by choosing $\theta^2 = \mu$. But if $\bar{\theta}^2 \neq \mu$, it is not true that $v^2(\theta^2) > v^2(\bar{\theta}^1)$. Hence in the Nash equilibrium sense, candidate 2 has no incentive to change from $\bar{\theta}^2$. A similar argument holds when $\bar{\theta}^2 = \mu$. Thus $\bar{\theta}^1 = \mu$ or $\bar{\theta}^2 = \mu$ implies that $(\bar{\theta}^1, \bar{\theta}^2)$ is a Nash equilibrium. □

The result in Theorem 8 may be seen as generalizing that of Wittman [1977], where (in a dynamic context) candidates choose their most preferred points from the set of those which defeat their incumbent opponents. Although there are problems with the implicit assumption that this is always possible, Wittman's results may be interpreted in a static form. In this interpretation, he shows (in his Proposition 5) that if a total median exists and if y^1 and y^2 are paired, that is, are exactly opposite μ from one another, then μ is a minimax strategy for either player against an opponent not at μ. That is, if $\theta^1 \neq \mu$ is the (immobile) incumbent then candidate 2 minimizes the plurality 1 can get if 2 chooses $\theta^2 = \mu$. The implication is that if 2 chooses μ, he will never have any incentive to change (under our assumptions); and candidate 1 may as well not change since he cannot defeat μ and can only tie it with $\theta^1 = \mu$ which is no better than having $\theta^2 = \mu$ win. Hence as in Theorem 8, $\bar{\theta}^1 = \mu$ or $\bar{\theta}^2 = \mu$ if and only if $(\bar{\theta}^1, \bar{\theta}^2)$ is a Nash equilibrium.

Our result with Type I preferences can be further extended to more general preference structures in the following sense: the sets C^1 and C^2, under the type I assumption, were just straight line segments from μ to y^1 and y^2. The same reasoning as before applies in general -- the winning candidate should always choose to be in, or arbitrarily close to, his C^j set. These may now be curves or even several curves, if the level curves of ϕ do not bound convex sets (note that the Lemma may not apply if voters' preferences are not Type I). In this setting, there may be a Nash equilibrium with neither candidate at μ. Suppose for some $\bar{\theta}^1$, $\bar{\theta}^2$ that $\bar{\theta}^1 \in \{C^1 \cap C^2\}$ but $\bar{\theta}^1 \neq \mu$; the simple assumption $\frac{y^1 - \mu}{\|y^1 - \mu\|} \neq \frac{y^2 - \mu}{\|y^2 - \mu\|}$ no longer prevents this. If $\bar{\theta}^2$ is on C^2 between y^2 and C^1 and if C^2 is a single curve, then candidate 2 may not wish to move. Since $\bar{\theta}^1$ is on C^2, candidate 2 cannot win without assuming a position θ^2 for which $v^2(\theta^2) < v^2(\bar{\theta}^1)$. Thus in general the opposition condition for having the equilibrium winning candidate at μ is much more complex than in the one-dimensional or Type I cases.

It is also plausible, of course, to suppose that candidates may have preferences over policy and prefer to win office or to win votes. In general this would be expressed as a utility function $v^i(\theta^1, \theta^2)$ for candidate i which takes the form

$$v^i(\theta^1, \theta^2) = \bar{v}^i(\phi(\theta^1, \theta^2), \theta^1, \theta^2) \quad \text{or}$$

$$v^i(\theta^1, \theta^2) = \bar{v}^i(V^i(\theta^1, \theta^2), \theta^1, \theta^2)$$

where V^i is the aggregate vote total for i and $\bar{v}^i$ depends only on the θ^i of the winning candidate. Equilibrium analysis could of course be performed on this model to study conditions for existence. Here we will simply note the effect of one simple form of combined preferences on equilibrium for policy-oriented candidates. Suppose candidate 1's preferences are represented by

$$w^1(\theta^1, \theta^2) = \begin{cases} v^1(\theta^1) + W_1 \text{ if candidate 1 wins} \\ v^1(\theta^2) \text{ if candidate 2 wins,} \end{cases}$$

where v^1 is the utility for policy outcomes, as before. Suppose α is a point in $C^1 \cap C^2$, either the total median μ or another intersection point. With the constant $W_1 > 0$, an extra satisfaction from winning office, (θ^1, α) will no longer be an equilibrium when $\phi(\theta^1, \alpha) < 0$; it is in candidate 1's interest to achieve a tie. In particular, if both candidates have $W_i > 0$, the only possible Nash equilibrium strategy pair is the convergent result (μ,μ).

3. UNCERTAIN ELECTION OUTCOMES

So far, candidates have been modeled as knowing for certain what the outcome of an election will be given both candidates' platform choices. However, almost any type of uncertainty which could be included in the spatial modeling framework would destroy this certainty, leaving candidates to achieve their goals in the face of a probabilistic outcome. Such a situation occurs whenever voter behavior has a random component if there are finitely many voters; this applies whether the behavior is really stochastic or is simply unpredictable to the candidate. Voters may be uncertain or inconsistent in their own preferences; or they may be unclear about the means-ends relationships between policies advocated, policies implemented, and desired social outcomes. Candidates may not know all about voters' preferences or beliefs, or even their aggregate distribution, so that the candidates would face uncertain election outcomes even if voters were really fully rational and had consistent preferences.

If the campaign is viewed as an information process in which candidates signal voters about platforms, another source of candidate uncertainty about outcomes appears. This is the case of Chapter 2, for the individual voter; "noise" may appear in candidates' signals or voters may err in evaluating the signals, so that each voter's final choice is stochastic. Voters' search behavior is determined partly by search costs and voters' priors as well. Thus to predict the outcome, candidates must know all about the subjective costs, priors, and sampling distributions of each voter. Given such detailed information requirements, it is certainly reasonable to expect candidates to be able to foresee the vote without considerable uncertainty.

The points about voters' decisions and the campaign information process made above might be applied in various combinations when electoral systems are modeled formally, each one being important in some cases and perhaps not in others. The important thing to notice is that each of these sources of uncertainty leads to a situation in which candidates, having chosen their campaign strategies, cannot predict the election outcome precisely. If candidates are rational they must behave as if the strategies define a whole set of possible outcomes together perhaps with an idea of the relative likelihoods of those outcomes.

Optimal behavior then may be entirely different from what it would be under uncertainty. Depending on the other assumptions made, conclusions about candidate strategy equilibria may also be affected.

4. OFFICE-ORIENTED CANDIDATES WITH IMPERFECT INFORMATION

Indeterminacies in voting at the individual level have been employed in the spatial modeling literature by Hinich and Ordeshook [1969, 1970] and Hinich, Ledyard, and Ordeshook [1972]. Here the stochastic part of the voting decision has been confined to the decision of whether or not to vote, expressed as a probability $g(x_\alpha - \theta^1, x_\alpha - \theta^2, \alpha)$ for individual α with bliss point x_α. The effect of this feature in the models has been essentially just to produce a tradeoff between gaining voters from the opponent and losing voters due to "alienation" (max $\{\|x_\alpha - \theta^1\|, \|x_\alpha - \theta^2\|\}$ gets too small) or "indifference" ($|\|x_\alpha - \theta^1\| - \|x_\alpha - \theta^2\||$ gets too small). As mentioned before, Hinich and Ordeshook [1970] found that sufficient sensitivity of turnout to alienation coupled with vote-maximizing behavior by candidates can produce divergent equilibria. The Hinich, Ledyard, and Ordeshook paper achieves a convergent Nash equilibrium result <u>without</u> requiring existence of a majority rule equilibrium. This is done by using concavity and convexity assumptions in a differential equation model of candidate response. In the Appendix to this chapter we will examine their approach in more detail, from a simpler candidate maximization viewpoint.

In these models, however, the problem facing the candidates is deterministic. Because there are infinitely many voters measured only by a probability distribution, the number of voters choosing each candidate and the number abstaining are completely determined by the values of θ^1 and θ^2. For example if $g(\theta^1, \theta^2, x)$ represents the probability of voting by a citizen with bliss point x, and $S^1 = \{x \mid \|x - \theta^1\| < \|x - \theta^2\|\}$ and $S^2 = \{x \mid \|x - \theta^1\| > \|x - \theta^2\|\}$, then the total vote achieved by candidates 1 and 2 might be expressed as

$$V^1(\theta^1, \theta^2) = \int_{S^1} f(x) g(\theta^1, \theta^2, x)dx \text{ and}$$

$$V^2(\theta^1, \theta^2) = \int_{S^2} f(x) g(\theta^1, \theta^2, x)dx,$$

respectively, with plurality for candidate 1

$$\phi(\theta^1, \theta^2) = V^1(\theta^1, \theta^2) - V^2(\theta^1, \theta^2).$$

The candidates' knowledge of individual voter behavior is not complete, but it is sufficient to predict the outcome of the election with certainty. We might still say that the candidates have perfect knowledge of the voters' policy preferences and of their perceptions of θ^1 and θ^2.

The stronger type of imperfect information occurs when the aggregate outcome cannot be predicted with certainty. If the candidates are assumed to have probabilistic information about outcomes, such a situation can be summed up in the form of probability-of-winning functions $P^1(\theta^1,\theta^2)$, $P^2(\theta^1,\theta^2)$ or as probability distributions of the proportion of votes to be received by each candidate $V^1(\theta^1, \theta^2)$ and $V^2(\theta^1, \theta^2)$. In this kind of model one might think, as before, of candidates who maximize expected vote or expected plurality, but now the expectations would be taken over V^1 and V^2 and the outcome would be really probabilistic. This stochastic feature could alter optimal candidate behavior in meaningful ways from that in the deterministic setting. Maximizing expected plurality of vote might no longer be an appropriate goal, depending on the form of the candidate's utility function for votes. The candidate may wish to maximize expected utility of plurality. In addition, the whole question of the rationality of the expected utility calculus becomes important, depending again on assumptions made about candidate preferences.

In the most straightforward interpretation of "office-oriented" candidate objectives, the candidate is interested only in whether he wins or loses. That is, he chooses θ^i to maximize the probability of winning in a constant sum game against the other candidate in which the payoffs are $P^1(\theta^1, \theta^2)$ and $P^2(\theta^1, \theta^2)$. Denzau and Kats [1977] have examined such a model, showing that under concavity assumptions on the individual level probabilities (they use a finite population), there exists a set of positions in which at least one candidate is assured of winning or tying the other. Although Denzau and Kats claim to be studying expected plurality maximization by the candidates, their solution sets

$Z_1 = \{\theta^1 \varepsilon S | \phi(\theta^1, \theta^2) \geq 0 \ \forall \ \theta^2\}$, etc., are winning (or tying) positions which have nothing to do with plurality maximization. However, we can use their definitions to show that at the individual level, a plurality maximizing position exists as well: let $g_j^i(u_j(\theta^1), u_j(\theta^2))$ be the (continuous) probability that individual j votes for candidate i when his utility for θ^1 is $u_j(\theta^1)$ and for θ^2, $u_j(\theta^2)$. Assume u_j is continuous and concave in θ^i, g_j^i concave in $u_j(\theta^i)$ and convex in its other argument. The total expected vote to be received by candidate 1 is $EV^1(\theta^1, \theta^2) = \sum_{j=1}^{N} g_j^1(u_j(\theta^1), u_j(\theta^2))$, a sum of concave functions in θ^1 and therefore itself concave in θ^1; likewise EV^1 is convex in θ^2. Similarly EV^2 is convex in θ^1 and concave in θ^2. The expected plurality for candidate 1,

$$E\phi^1(\theta^1, \theta^2) = EV^1(\theta^1, \theta^2) - EV^2(\theta^1, \theta^2)$$

is concave in θ^1, and $E\phi^2$ is concave in θ^2. Existence of an equilibrium strategy pair follows from a simple fixed-point argument.

Theorem 9: Let S be a compact issue space. Under the assumptions above, there exists a Nash equilibrium strategy pair $(\bar{\theta}^1, \bar{\theta}^2)$.

Proof: Let $r_1(\theta^2)$ be the position θ^1 such that

$$E\phi^1(\theta^1, \theta^2) = \max E\phi^1(\theta, \theta^2) .$$

Existence and uniqueness of $r_1(\theta^2)$ are guaranteed by concavity of $E\phi^1$ in θ^1. Similarly define $r_2(\theta^1)$. Let $R : S \times S \to S \times S$ be a function mapping each (θ^1, θ^2) into $(r_1(\theta^2), r_2(\theta^1))$. Since p^i_j and u_j are continuous, $E\phi^i$ must be continuous, and a simple argument shows r_1 and r_2 continuous as well. By the Brouwer fixed point theorem, then, R must have a fixed point $(\bar{\theta}^1, \bar{\theta}^2)$ such that $\bar{\theta}^1 = r_1(\bar{\theta}^2)$ and $\bar{\theta}^2 = r_2(\bar{\theta}^1)$; that is, if one candidate keeps his position, the other will not desire to move either. This is the Nash equilibrium strategy pair. □

Strict concavity can also be dispensed with; the Kakutani fixed-point theorem can be applied to guarantee a set of Nash equilibria when piecewise linearity is allowed in g^i_j and u_j.

The analysis which does appear in Denzau and Kats does not guarantee an equilibrium for candidates who maximize their probabilities of winning. If Z_1 and Z_2 are both nonempty, there may be majority rule cycles. Existence of equilibrium requires a niceness condition on the optimization of P^1 and P^2, such as concavity, under which an argument similar to that in Theorem 9 can be used. There is a result due to Wittman [1975] showing that if equilibrium does exist, it must be true that $\bar{\theta}^1 = \bar{\theta}^2$. Wittman uses first- and second-order conditions on differentiable individual probability functions to show that (under conditions of symmetry between the candidates in the individual voting decision, of course) simultaneous maximization of the probabilities of winning by two candidates implies that they take the same position. Wittman also employs an assumption similar to concavity on g_j^i.

5. POLICY-ORIENTED CANDIDATES WITH IMPERFECT INFORMATION

It is perhaps surprising in light of the results in section 2 and the results of Wittman in section 4 that equilibrium in the case of policy-oriented candidates facing uncertain aggregate outcomes <u>precludes</u> convergent strategies by the candidates. Here the necessary condition of "opposition" of candidate preferences is simply that their bliss points not be identical.

Suppose that $P^1(\theta^1, \theta^2)$ is the probability that candidate i wins when positions θ^1 and θ^2 are taken, and that the candidates wish to maximize the expected utility of the policy outcome, $P^1(\theta^1, \theta^2)v^i(\theta^1) + P^2(\theta^1, \theta^2)v^i(\theta^2)$. Assume that the v^i are not flat, i.e., that $\nabla v^i(\theta) = 0 \Rightarrow \nabla v^i(\theta + \varepsilon) \neq 0$ for all $\varepsilon \in \mathbb{R}^n$ such that $\|\varepsilon\|$ is sufficiently small.

<u>Theorem 10</u>: Let α be a point in the interior of S, which is not the bliss point of both candidates. Suppose $P^1 > 0$ and $P^2 > 0$ on some open set containing α. Then (α,α) is not an equilibrium strategy pair.

Proof: Suppose $\theta^2 = \alpha$. For some $\delta \in \mathbb{R}^m$ with $\|\delta\|$ sufficiently small, $v^1(\alpha + \delta) > v^1(\alpha)$. For some sufficiently small $\varepsilon > 0$, $P^1(\alpha + \varepsilon\delta, \theta^2) > 0$. At α, candidate 1's expected utility would be

$$P^1(\alpha,\alpha)v^1(\alpha) + P^2(\alpha,\alpha)v^1(\alpha) = v^1(\alpha),$$

while at $\alpha + \varepsilon\delta$ it would be

$$P^1(\alpha + \varepsilon\delta, \alpha)v^1(\alpha + \varepsilon\delta) + P^2(\alpha + \varepsilon\delta, \alpha)v^1(\alpha).$$

Since $P^1(\alpha + \varepsilon\delta, \alpha) > 0$ and $v^1(\alpha + \varepsilon\delta) > v^1(\alpha)$, $\theta^1 = \alpha + \varepsilon\delta$ is preferred by candidate 1 to α if $\theta^2 = \alpha$. Thus (α, α) is not a Nash equilibrium. □

Under maximization of expected utility of the policy outcome, a Nash equilibrium strategy pair is a pair $(\bar{\theta}^1, \bar{\theta}^2)$ such that $\bar{\theta}^1$ maximizes

$$P^1(\theta, \bar{\theta}^2)v^1(\theta) + P^2(\theta, \bar{\theta}^2)v^1(\bar{\theta}^2)$$
$$= P^1(\theta, \bar{\theta}^2)[v^1(\theta) - v^1(\bar{\theta}^2)] + v^1(\bar{\theta}^2),$$

(the latter since $P^1 + P^2 = 1$), and $\bar{\theta}^2$ maximizes

$$P^1(\bar{\theta}^1, \theta)v^2(\bar{\theta}^1) + P^2(\bar{\theta}^1, \theta)v^2(\theta)$$
$$= P^1(\bar{\theta}^1, \theta)[v^2(\bar{\theta}^1) - v^2(\theta)] + v^2(\theta).$$

Thus conditions for existence of such an equilibrium hinge on the shapes of <u>products</u> of probability functions and utility functions. Still, equilibrium can exist. Suppose for example that the probability functions are fairly flat while the utility functions are both strongly concave; then $P^1v^i(\theta^1, \theta^2) + (1 - P^1)v^i(\theta^1, \theta^2)$ is likely to be a nice concave function for $i = 1,2$, in which case continuity and compactness can be applied in a fixed-point argument for existence.

An important point to be made about policy-oriented candidates facing probabilistic outcomes is that even if equilibrium does exist, if the outcome is symmetric between the two candidates ($P^1(\theta^1, \theta^2) = P^2(\theta^2, \theta^1)$), and if the voters are distributed so as to give nice, symmetric probability-of-winning functions (analogous to existence of a total median), it is still impossible for a convergent equilibrium to result. In particular suppose there is a point μ such that the candidate closer to μ always has the better chance of winning. Each candidate has an incentive to move closer to μ than his opponent will be, but the tendency to stay close to their own preferred points offsets this. Thus the possibility of winning even with a position further from μ changes the situation rather drastically from that examined in section 2, where candidates had perfect information about electoral outcomes.

If we imagine a continuum of information conditions, from perfect information to total ignorance, we can see more completely the effects of uncertainty on the platforms chosen by the candidates. Suppose a total median μ exists and that the probability functions behave as above -- being closer to μ means a larger chance of winning. Under complete certainty

$$P^1(\theta^1, \theta^2) = \begin{cases} 1 \text{ if } \|\theta^1 - \mu\| < \|\theta^2 - \mu\|, \\ 0 \text{ if } \|\theta^1 - \mu\| > \|\theta^2 - \mu\|, \text{ and} \\ 1/2 \text{ if } \|\theta^1 - \mu\| = \|\theta^2 - \mu\| \end{cases}$$

and as in section 2 the candidates converge to (μ,μ) under opposition conditions. With slight uncertainty however, the jump discontinuity in P^1 between $\|\theta^1 - \mu\| < \|\theta^2 - \mu\|$ and $\|\theta^1 - \mu\| > \|\theta^2 - \mu\|$ is smoothed out; there is a neighborhood of μ where $1 > P^1(\theta^1, \mu) > 0$, and Proposition 3 can be applied to show that (μ, μ) is no longer an equilibrium. It is easy to see, in fact, that the effect of this smoothing would be to send the equilibrium strategies, if they exist, along each player's utility gradient at μ. Suppose that more uncertainty is added to the model via a "mode-preserving spread" of the probability functions: that is, for any fixed $\bar{\theta}^2$ let values of P^1 near the $P^1(\theta^1, \bar{\theta}^2)$-maximizing value of θ^1 (namely $\mu = \theta^1$) get smaller while the tails of $P^1(\theta^1, \bar{\theta}^2)$ get thicker in such a way as to preserve μ as the mode of P^1 (note that P^1 is <u>not</u> a probability measure on S x S so that there is no question of preserving a particular value of its integral over the whole space). If existence of equilibrium is retained, it is clear that candidate 1's optimal platform moves closer to his bliss point, because for each value of $\bar{\theta}^2$, values of $P^1(\theta^1, \bar{\theta}^2)v^1(\theta^1) + (1 - P^1)v^1(\bar{\theta}^2)$ near $\theta^1 = \mu$ get smaller while values further away get larger (we assume of course that θ^1 will be chosen so that $v^1(\theta^1) > v^2(\bar{\theta}^2)$). Thus the more uncertainty there is about the election outcome, the more the platforms will diverge from the total median or other central point of P^1 and P^2.

Finally, we can also consider the combination of office- and policy-orientation in this context. Suppose candidates' preferences are as in section 2,

$$w^1(\theta^1, \theta^2) = \begin{cases} v^1(\theta^1) + W_1 & \text{if candidate 1 wins} \\ v^1(\theta^2) & \text{if candidate 2 wins.} \end{cases}$$

Then the expected value of the election outcome to candidate 1 is

$$P^1(\theta^1, \theta^2)[v^1(\theta^1) + W_1] + [1 - P^1(\theta^1, \theta^2)]v^1(\theta^2).$$

It is easy to see that the proof of Proposition 3 goes through exactly as before; the addition of the term $W_1 > 0$ lessens the incentive to move away from a convergent strategy pair but does not eliminate it. The proof of Theorem 10 may still apply; convergence now hinges on the candidate's comparison of his gain from the term $P^1(\alpha + \varepsilon\delta, \alpha)v^1(\alpha + \varepsilon\delta)$ with his loss from the term $P^1(\alpha + \varepsilon\delta, \alpha)W_1$. The tendency away from convergence is thus weakened but not eliminated.

6. CONCLUSION

Examining four basic combinations of candidates' goals and their certainty about election outcomes we have noted various implications for the kinds of equilibrium strategies which can occur. Most important is the case of policy-oriented candidates with uncertainty, where it is impossible for $(\theta^1, \theta^2) = (\alpha, \alpha)$ to be an equilibrium strategy for any α, even if α is the total median. Considering mixed candidate motivations we have seen that this result still holds. We can conclude that for a broad class of spatial models, any degree of policy orientation in candidates' preferences along with any degree of "smooth" uncertainty (e.g. characterized by probability-of-winning functions which are continuous and positive on relevant areas of issue space) carries pressure toward divergence of optimal candidate strategies. This tendency is independent of asymmetries between the candidates, absence of equilibria or multiple equilibria, or other conditions which clearly might also result in divergence of candidate platforms from one another. Put another way, if we believe that policy considerations enter into candidates' preferences then the presence of uncertainty alone determines whether it is possible in general for convergence to the median voter to represent an equilibrium outcome.

Appendix

CONCAVITY AND EQUILIBRIUM IN SPATIAL MODELS

As discussed in Section 4 of this chapter, Hinich et al. [1972] used various concavity and convexity assumptions to prove existence of an equilibrium strategy pair for two candidates. Although theirs is actually a dynamic, differential equation model of candidate adjustment, its salient features can be understood through a simple static model in which a fixed-point theorem is applied. In this Appendix we develop such an approach to their model. Then we show that, if voters are assumed to behave as those in Chapter 2, the convexity and concavity assumptions in Hinich et al. can be replaced by a single concavity assumption on the voters' sampling distributions.

Let (A,λ) be a measurable space which is an index set of citizens $\alpha \in A$. Let S be the (compact, convex) issue space. Each citizen is assumed by Hinich et al. to have a utility function $u(\theta,\alpha)$ defined over SxA which is concave in θ. Let θ^1 be the position of candidate 1 in S and θ^2 the position of candidate 2. Hinich et al. assign to each $\alpha \in A$ a probability function $g^1(u(\theta^1,\alpha), u(\theta^2,\alpha),\alpha)$ of voting for candidate 1 when the candidates are at θ^1 and θ^2, and $g^2(u(\theta^1,\alpha), u(\theta^2,\alpha),\alpha)$ of voting for candidate 2. They assume that, for all $\alpha \in A$, g^1 is concave in its first argument and convex in its second, and that g^2 is convex in its first argument and concave in its second. Let us define $p^i(\theta^1,\theta^2,\alpha) = g^i(u(\theta^1,\alpha), u(\theta^2,\alpha),\alpha)$ for $i = 1,2$; then the effect (and the goal) of the above concavity and convexity assumptions is to make p^1 concave in θ^1 and convex in θ^2, and vice-versa for p^2. Now, adding and subtracting to get the plurality function preserves concavity and convexity in the appropriate arguments: the total expected vote for candidate i is

$$V^i(\theta^1,\theta^2) = \int_A p^i(\theta^1,\theta^2,\alpha)d\lambda,$$

and plurality for candidate 1 over candidate 2 is just $\phi = V^1 - V^2$. This function is to be maximized by candidate 1 and minimized by candidate 2, and it is appropriately concave in θ^1 and convex in θ^2. Hinich et al. make the additional assumption that strict concavity or strict convexity hold for u, g^1, and g^2 on a set of positive measure in A; this assures strict concavity and convexity of ϕ. Now, just as in the proof of Theorem 9, we can define continuous reaction functions for the two candidates and apply Brouwer's fixed point theorem to get existence of a candidate equilibrium.

In Chapter 2 we noted that, given a sampling distribution for single observations, a distribution can be defined on the infinite sequences of those observations. Accordingly, let $h(x|\theta^1,\theta^2,\alpha)$ be the probability density of the infinite sequence x of pairs of observations on candidate 1 and candidate 2 when θ^1 and θ^2 are the candidates' true positions, for voter $\alpha \in A$. In place of the assumptions of Hinich et al., we assume that h is concave in θ^1 and convex in θ^2. In terms of the first argument of h, this could be obtained alternatively by assuming $h(x|\overline{\theta},\alpha) = h(x-\overline{\theta}|0,\alpha)$ and assuming h concave in x^1 (the observations on candidate 1) and convex in x^2, where $\overline{\theta} = (\theta^1,\theta^2)$. Furthermore, we assume that for each α in some set of positive λ-measure in A, there is a subset of sequences of observations x having positive probability on which h is strictly concave in θ^1 and strictly convex in θ^2. (This subset of observations may be different for different α.) Let Ω^1 be the set of sequences of observations such that the optimal sampling plan says to stop and choose candidate 1 after finitely many observations; we assume Ω^1 is integrable. Then we can define the same p^1 function as before by

$$p^1(\theta^1,\theta^2,\alpha) = \int_{\Omega^1} h(x|\theta^1,\theta^2,\alpha)\,dx.$$

Exactly as before, now, p^1 is strictly concave in θ^1 and strictly convex in θ^2 on a set of positive λ-measure in A; we can proceed exactly as before to prove existence of a candidate equilibrium.

Thus instead of making assumptions about the shapes of u, g^1, and g^2, we could also get the equilibrium result by making assumptions about h. This is somewhat more appealing since it may not even be reasonable to assume concavity of u on an issue space; on the other hand, making assumptions about h is not the same as making them about the simple sampling distribution, products of which describe the finite-dimensional marginals of h. Since a product of concave functions is not generally concave, we cannot back up to make more understandable assumptions on the single-observation density function.

REFERENCES

Cohen Linda. "Cyclic Sets in Multidimensional Voting Models." Social Science Working Paper No. 172. Pasadena, Calif.: California Institute of Technology, 1977.

Cohen, Linda and Steven Matthews. "Constrained Plott Equilibria, Directional Equilibria, and Global Cycling Sets." Social Science Working Paper No. 178. Pasadena, Calif.: California Institute of Technology, 1977.

Davis, Otto A. and Melvin J. Hinich. "A Mathematical Model of Policy Formation in a Democratic Society." In Mathematical Applications in Political Science II, edited by Joseph Bernd. Dallas: Southern Methodist University Press, 1966.

____________________. "Some Results Related to a Mathematical Model of Policy Formation in a Democratic Society." In Mathematical Applications in Political Science III, edited by Joseph Bernd. Charlottesville, Va.: University Press of Virginia, 1967.

____________________. "On the Power and Importance of the Mean Preference in a Mathematical Model of Democratic Choice." Public Choice 5 (Fall 1968):59-72.

Davis, Otto , Melvin J. Hinich, and Peter C. Ordeshook. "An Expository Development of a Mathematical Model of the Electoral Process." American Political Science Review 64 (1970):426-448.

Denzau, Arthur T. and Amoz Kats. "Expected Plurality Voting Equilibrium and Social Choice Functions." Review of Economic Studies 44 (1977):227-233.

Downs, Anthony. An Economic Theory of Democracy. New York: Harper and Brothers, 1957.

Hinich, Melvin J., John O. Ledyard, and Peter C. Ordeshook. "Nonvoting and the Existence of Equilibrium under Majority Rule." Journal of Economic Theory 4 (1972):144-153.

Hinich, Melvin J. and Peter C. Ordeshook. "Abstentions and Equilibrium in the Electoral Process." Public Choice 7 (Fall 1969):81-106.

____________________. "Plurality Maximization Vs. Vote Maximization: A Spatial Analysis with Variable Participation." American Political Science Review 64 (1970):772-791.

Hotelling, Harold. "Stability in Competition." The Economic Journal 39 (1929):41-57.

Hoyer, R. W. and Lawrence S. Mayer. "Comparing Strategies in a Spatial Model of Electoral Competition." American Journal of Political Science 18 (1974):501-523.

Kramer, Gerald H. "A Dynamical Model of Political Equilibrium." Cowles Foundation Discussion Paper No. 396. New Haven: Yale University, 1975.

________________. "Some Extensions of a Dynamical Model of Electoral Competition." Mimeographed. New Haven: Yale University, 1977.

McKelvey, Richard D. "Intransitivities in Multidimensional Voting Models and Some Implications for Agenda Control." Journal of Economic Theory 12 (1976):472-482.

________________. "General Conditions for Global Intransitivities in Formal Voting Model." Mimeographed. Pittsburgh: Carnegie-Mellon University, 1977.

Plott, Charles R. "A Notion of Equilibrium and Its Possibility under Majority Rule." American Economic Review 57 (1967): 787-806.

Schofield, Norman. "Generic Instability of Voting Games." Mimeographed. Austin: University of Texas at Austin, 1978.

Shepsle, Kenneth A. "Institutional Arrangements and Equilibrium in Multidimensional Voting Models." Mimeographed. St. Louis, Mo.: Washington University, 1977.

__________________. "Institutional Structure and Policy Choice: Some Comparative Statics of Amendment Control Procedures." Mimeographed. St. Louis, Mo.: Washington University, 1978.

Stokes, Donald E. "Spatial Models of Party Competition." American Political Science Review 57 (1963):368-377. Also in Elections and the Political Order, edited by Angus Campbell, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. New York: John Wiley and Sons, 1966.

Wittman, Donald. "Candidate Equilibrium." Mimeographed. Santa Cruz, Calif.: University of California, Santa Cruz, 1975.

_______________. "Candidates with Policy Preferences: A Dynamic Model." Journal of Economic Theory 14 (1977):180-189.

Chapter 5

CONCLUSIONS

In this study, we have attempted to learn more about how an electoral system functions in the presence of ignorance and uncertainty by examining the behavior of rational voters and candidates under those conditions. We have explicitly modelled the voter as having incomplete information and facing barriers to obtaining more, barriers of opportunity cost and of "noise" in observations. Candidates, partly as a result of the voters' situation, must determine their election strategies without knowing exactly what the results of their choices may be. What does this theoretical study tell us about democratic government and politics?

Most important, perhaps, is the voter model itself. It represents a model of full voter rationality, subsuming both the voting decision under uncertainty and the process of alleviating that uncertainty. This rationality is summarized by the functional equation of Chapter 2, describing how the gathering of information depends upon costs, previous information, and the further information available. The functional equation has a simple heuristic interpretation: it represents the voter asking himself whether it is "worth it" to take another observation, exactly as microeconomic consumer theory's constrained utility maximization describes a consumer asking himself, at the margin, whether another unit of a given commodity is worth its opportunity cost in foregoing some alternative commodity. Moreover, Theorems 1 and 2 demonstrate that under fairly general conditions the functional equation constitutes a consistent, well-defined rule for rational decision-making. Thus it makes sense to talk about rational voting under costly and imperfect information.

Using the functional equation, we were able to demonstrate some of the characteristics of this rational voting process. Under some priors, Theorem 3 shows that no sampling need take place no matter how low the cost; Theorem 4 demonstrates that under other conditions there is always a positive level of cost low enough that sampling will take place. Taken together, these theorems demonstrate the sense in which an increase in search cost brings about a decrease in sampling. Theorem 5 then relates changes in risk to the value of search as developed in Theorems 3 and 4, showing that if the riskiness of the expected value of stopping after another observation increases, more sampling will occur. This result is independent of the risk preference of the voter. Unfortunately no corresponding result could be obtained for risk changes in the prior or sampling distributions themselves; in particular, it might not be true that a mean-preserving spread in the prior or in the sampling distribution would lead to a mean-preserving spread in the expectation of the value of stopping after the next observation.

Under very strong regularity conditions on preferences and on the subjective distributions we were able to relate, in Theorem 6, the sampling process to the final choice. In particular, if the final observation, taken when priors are uniform, favors one candidate, then that candidate will have the highest posterior expected value and thus will be the voter's choice. In Theorem 7, when the voter operates not with perceptions of candidate positions but with utility levels only, we showed that the restrictive conditions of Theorem 6 can be relaxed considerably. However, the counterexamples following Theorem 6 show that this is not the case in general: relaxation of any single condition causes the theorem to fail. These counterexamples demonstrate how unpredictable voter behavior can be when the actual issue positions of the candidates are observed. The issue space model invokes many considerations of the riskiness and consistency of the candidates' signals and the steepness of the utility function which do not appear in the case of direct observation of utility levels. On the one hand, this is highly intuitive: the voter who has no concern with actual issue positions has, a fortiori, no concern with how many different ways a candidate may claim to be able to achieve a given utility level; but the voter who perceives the candidates in an issue space is implicitly putting to use his knowledge about how platforms connect with policies and outcomes. On the other hand, seemingly unnatural results may spring up in the latter case. For example, outlying observations in opposite directions from the voter's bliss point may strengthen the voter's expectation that the candidate is in reality near the the bliss point. The counterexamples thus help demonstrate that the two models have drastically different properties, although even such sophisticated analyses as those of Powell [1974] and Fiorina [1977] implicitly regard them as being interchangeable.

The properties of our model of rational voting have important theoretical implications for the operation of the electoral system. In Chapter 4 we saw that voter uncertainty contributes to candidate uncertainty which may, depending on candidate goals, influence the candidates to not adopt convergent platforms near the median voter. Outside the purview of Chapter 4 lie more direct conclusions. First, the voter model suggests that candidates should view voters as they appear in the model and not as simple utility maximizers without an information-gathering problem. To campaign effectively, candidates should learn about voters' priors, their subjective sampling distributions, and the search costs they face. Second, the voter model suggests strongly that the reliability and availability of information for the voters is of as basic importance in a democracy as the right to vote itself. The issue is whether the people influence government policies; to do this they require not only an electoral process but the information to use it. Certainly this has been realized before in various forms, but the voting model of this study makes paramount the effects of costs and reliability of information on the relationship between the real wishes of a voter and his choice of candidates. Finally, we can relate this study to Schattschneider's [1960] problem, whether ordinary citizens can be served by a democracy. In the terms of our model, if government is to be responsive then citizens must have sufficiently low costs of information, in the form of education, political sophistication, and interest in politics. In addition, the information generated about the candidates must be generated in keeping with the voters' expectations about that information; that is, voters' subjective sampling distributions must resemble the "actual" sampling distribution. Only under these conditions will voters tend to vote for the candidate who best represents their needs and wishes. Then, if electoral pressures are sufficient to influence candidates (as is the case in all four models of Chapter 4), the voters' wishes will be reflected, to some degree, in government policy. The voting model of this study shows finally that, in principle, there can be rational voting; whether this will be sufficiently related to policy thus depends on the realities represented by the model's parameters of costs and expectations.

The voting model has important implications for the empirical study of voting behavior as well. Most obviously, the model demonstrates that rational voting, even if it is basically issue related, is compatible with the observation of a paucity of detailed political information among voters, the use of very rough cues for voting choice such as candidate personalities and party labels, and the persistence of party identification in the face of strong short-term forces. The apparent differences between countries in the importance of party identification (see Budge et al. [1976]) finds a ready explanation in terms of the varying value of party as an informational cue relative to other possible cues. This explanation suggests obvious possible tests of its validity, such as examining in more detail how voters see the connection between party and policy, the strength and stability of this perceived connection, and the connections between policy and non-party cues. Both survey data and contextual data would be important for such tests.

The voting model itself is amenable to testing by combining the direct tests suggested by Fiorina's [1977] model with consideration of voters' subjective notions of information reliability and cost; besides proxies such as education and media use, new survey items would be needed to explore the reliability which voters attach to their own beliefs and to candidates' claims. Using the findings that would result from these tests, empirical researchers could then attempt to predict voting behavior and the effect of political campaigns by explicitly taking account of the voters' information-gathering process.

Future research along the lines of this study must address the candidates' side of the electoral system. Given voters who behave like those described here, what are the incentives for candidates? It is important to determine whether candidates will tend to provide information in such a way that voters' subjective sampling distributions will be realistic, if the candidates know the properties of those distributions. Also, we must examine the nature of the candidate's problem in choosing between different kinds of information to provide, information on different issues, through different channels, and so on. How these candidate problems should be modelled in the first place is perhaps the most difficult question, because an appropriate simplification such as the search model for voters is not so readily apparent for candidates. It may even be that an appropriate model of candidate behavior requires a version of "information" not compatible with that found in the present voter search model.

What the present study suggests, at least, is that any model of rational voting must take into account the problem of information acquisition by the voter if the observed phenomena of voting behavior are to be explained, and if the connection between voting and issues is to be understood. Yet imperfect information cannot be invoked in cavalier fashion to explain discrepancies in rational choice models. As we have seen, even intuitively "obvious" ideas may require difficult proof in a model which treats imperfect information explicitly; and many plausible hypotheses turn out to be false in general (consider again the counterexamples of Chapter 2), so that the phenomena to be explained require not just imperfect information but specific parameter values in the model as well. Therefore, this study represents a step forward in the formal modelling of voting behavior, and an example of what imperfect information in political behavior really entails.

REFERENCES

Budge, Ian, Ivor Crewe, and Dennis Farlie, eds. Party Identification and Beyond. London: John Wiley and Sons, 1976.

Fiorina, Morris P. "An Outline for a Model of Party Choice." American Journal of Political Science 21 (1977):601-626.

Powell, Lynda Watts. "A Bayesian Model of Individual Voting Decisions in Two-Candidate Elections." Mimeographed. Albany, N.Y.: Graduate School of Public Affairs, State University of New York at Albany, 1974. (Forthcoming in Journal of Mathematical Sociology 6 (1978).)

Schattschneider, E.E. The Semisovereign People. New York: Holt, Rinehart, and Winston, 1960.