An Entropy Measure of Uncertainty in Vote Choice

Jeff Gill

University of California, Davis

Abstract

We examine voters' uncertainty as they assess candidates' policy positions in the 1994 congressional election and test the hypothesis that the *Contract with America* reduced voter uncertainty about the issue positions of Republican House candidates. This is done with an aggregate evaluation of issue uncertainty and corresponding vote choice where the uncertainty parameterization is derived from an entropy calculation on a set of salient election issues. The primary advantage is that it requires very few assumptions about the nature of the data. The entropic model suggests that voters used the written and explicit Republican agenda as a means of reducing issue uncertainty without substantially increasing time spent evaluating candidate positions.

Introduction: Information, Uncertainty, and Voting

Complexity, obfuscation, vagueness, and uncertainty are permanent features of American electoral politics (Aldrich, *et al.* 1982). As modern social and economic issues become more complicated and legislatures increasingly dodge controversial and divisive votes, a greater number of poorly explained and widely misunderstood issues are thrust onto the ballot (Gafke and Leuthod 1979). Complex or lengthy ballot measures can also confuse voters (Gerber 1996). These are commonly worded in stern legal language and many voters lack the time or analytical background necessary to determine policy consequences. For example, the 1996 ballot initiative in California to eliminate state supported affirmative action programs, Proposition 209, confused many voters on its actual intent (Los Angeles Times, October 31, 1996, p.A22). The words "affirmative action" do not appear anywhere in the bill's language, and the bill is disingenuously entitled the "California Civil Rights Act."

Candidates often have motivations to make their issue positions deliberately vague (Downs 1957; Franklin 1991; Glazer 1990; Shepsle 1972). This can be out of a desire not to alienate certain constituency groups or simply that the candidate lacks a strongly defined position. There is evidence of this phenomenon even in very high salience elections. Bob Dole made deliberately vague statements on his abortion position in the 1996 presidential election after conservatives in the Republican party successfully placed strong anti-abortion rights language into the party platform at the San Diego convention (San Francisco Chronicle, August 5, 1996, p.A1). In campaigns it is a well-worn strategy to recast the language of an opponent's issue position into something less popular with voters (Nimmo 1970), and with frequent iterations of this practice has a generally confusing effect on the electorate.

So why do many voters endure high levels of uncertainty and cast their ballot? The answer is elusive, though well studied in the literature (Downs 1957;

Ferejohn & Fiorina 1974; Olson 1965; Wolfinger & Rosenstone 1980; Abramson & Aldrich 1982; Burnham 1987; Palfrey and Rosenthal 1985; Teixera 1992). Feddersen and Pesendorfer (1996, 1997, 1999) find that "asymmetric information" critically changes the calculus of voting by sufficiently blurring the distinctions between candidates for some citizens such that they are better off *not* voting at all. This work gives two findings of interest here. First, lower levels of ("private") information possessed by voters affect their ability or willingness to use perceived issue distance to sort-out preferred vote choices. Second, the observation that many vote, but some do not, implies that levels uncertainty must vary across individuals.

Rather than address the effect of uncertainty on the decision to turnout, this work evaluates the effect of uncertainty on the vote choice of those that *do* turnout. This focus stems from the observation that vote choice models developed without incorporating a voter uncertainty component miss an important criterion of the vote choice: "Models that rely on the assumption of complete information may offer a misleading view of the democratic process." (Ordeshook 1986, p.187).

While considerable attention has been paid to voting under uncertainty in two-candidate elections where voters abstain due to indifference or ignorance, we still do not have a complete theory leading to equilibrium conditions (Coughlin 1990) across a more complete range of circumstances. Furthermore, Hinich (1977) found that even a small amount of uncertainty is enough to destroy Black's elegant medium voter equilibrium result. Bartels (1986) used the "Don't Know" response as a measure of uncertainty, but this does not indicate the level of uncertainty for respondent provided responses. Also, Coombs and Coombs (1976) identify a difference in "Don't Know" answers: researcher created scale dependencies, and actual issue ambiguity. Furthermore, this issue ambiguity can be from ignorance, uncertainty, or just indecision (Sanchez and Morchio 1992).

Alvarez and Franklin (1994) note the additional problem that there is not a

consensus in the literature on how uncertainty is measured, although they were able to design a survey that provided very specific information on issue-level uncertainty to reveal subtle patterns in the form of the responses. More recently, Alvarez (1998) takes a comprehensive look at this question focusing on presidential elections. He finds that voters do, in fact, seek and acquire sufficient information to make informed vote decisions, at least at the presidential level.

Bartels (1988) studies the 1980 presidential election and finds that voters in general dislike uncertainty and this uncertainty, measured as variance around candidate positions, can exceed perceived spatial distances between the candidates. Like the Bartels uncertainty model, the approach developed here rests on the work of Enelow & Hinich (1981) and Shepsle (1972) in that issue positions are not singular and fixed. Instead they are represented by probability distribution functions with manifest observations (samples). It is precisely the quality of these samples, which provide information about the underlying true distribution of candidate positions, that affects levels of voter certainty. Such a theoretical model leads directly to a statistical framework where the mean is used as a point estimates of issue position and the variance represents uncertainty.

Objectives

The approach developed here leverages a unique aspect of the 1994 midterm election to evaluate how voters interpret differing levels of information about House candidate positions. In this election the Republican leadership in the House developed a well-defined and explicit issue-agenda. Republican incumbents and challengers signed the *Contract with America* promising that if elected to a majority they would initiate eight congressional reforms and bring ten specified issues (eventually requiring many more actual bills) to the House floor for a vote within the first 100 days of the 104th Congress. The voting behavior model here focuses on the uncertainty difference between Republican and Democratic candidates in this election, and asks whether there exist differing levels of information uncer-

3

tainty between the party that controlled the House of Representatives for forty years, and a minority party making explicit electoral promises.

Unlike previous measures of uncertainty, which come from individual survey based indicators or respondent self-assessments, the entropy form of uncertainty uses measures of respondent uncertainty that are produced by the format and content of the survey as assessed by the full sample. For purely methodological reasons researchers have been forced into the model assumption that all people have the exact same indifference to uncertainty and that this applies to all candidates evaluated. By replacing the variance term with an entropy term in the respondent's utility function, uncertainty is tied to the structure of the question choice and the full sample evaluation of the particular issue complexity. Therefore, since it is unrealistic to assume homogeneity of uncertainty threshold in respondents, the gain in moving away from the statistical variance term as a measure of issue uncertainty to an entropy measure of issue uncertainty can be substantial.

A Spatial Model of Vote Choice and Uncertainty

In this section a model of candidate vote choice is developed using the entropy concept as a relative measure of voter uncertainty between Democratic and Republican congressional candidates. This model is based on the well-known proximity spatial model developed by Downs (1957), Davis, Hinich, and Ordeshook (1970), Hinich and Pollard (1981), Enelow and Hinich (1981, 1984a, 1984b, 1990), and others. Like many works, the model relies on the assumption that voters seek a candidate choice that minimizes the squared Euclidean distance between their own issue preference and the perceived position of the candidate. However, issue distance is not a complete explanation of voter preferences and empirical models generally need to incorporate non-policy and attitudinal covariates as well.

Define a utility function, U_{ij}, that specifies the value of candidate j to voter i such that utility reduces for increasing distance between the perceived candi-

4

date position on issues and the respondents own position on these issues, and for increasing uncertainty about placing the candidates positions. So voters dislike candidates who are both further from their preferred policy positions and less clear on where they stand on these issues. Obviously we would also want to include terms that estimate the effects from non-issue contributions such as party identification, political interest, media exposure, and demographics. If we assume that the assessment of issue difference is squared, then this utility function for the ith respondent and the jth takes the form:

$$U_{ij} = \sum_{k=1}^{K} \left[-(C_{ijk} - R_{ik})^2 \right] + O_{ij} \tag{1}$$

where C_{ijk} is respondent i's placement of candidate j on policy issue k, R_{ik} is respondent i's self-placement on issue k, and O_{ij} is the effect of non-issue factors that contribute to respondent i's utility for candidate j. Voters are assumed to see candidates' issue positions at any moment in time as imprecise approximations due to candidate personality, campaign, and psychological effects, so we can therefore model this perception as realizations from a probability distribution (Bartels 1988; Enelow and Hinich 1981; Shepsle 1972). Now U_{ij} is also a realization from some underlying probability distribution, and voters will make decisions based on their *expected* utility (Grafstein 1991). This produces, with minor algebra (Bartels 1988, p.711), an expression for the expected utility of candidate j to voter i:

$$\text{E}[U_{ij}] = \sum_{k=1}^{K} \left[-(\text{E}[C_{ijk}] - R_{ik})^2 - \text{Var}[C_{ijk}] \right] + O_{ij} \tag{2}$$

where $\text{E}[C_{ijk}]$ is the expected position of the candidate and $\text{Var}[C_{ijk}]$ and the variance of the voters perception, both for issue k. So voters will prefer a candidate with a higher expected utility subject to the possibility of making an error in placing the candidate's issue positions.

The methodological problem with (2) is that the $\text{Var}[C_{ijk}]$ term cannot be directly observed except with specialized panel data. Bartels finesses this problem by assuming that uncertainty is a function of observable candidate and respondent characteristics provided in the survey data. He then posits a variance threshold (constant across candidates, issues, and respondents) which defines when respondents will make placements, and estimates from a linear model a quantity that is proportion to the probability that the respondent refuses to make a placement. Unfortunately this approach requires the assumption of homogeneity of the uncertainty (variance) threshold, and leads to a more complex, two-stage estimation procedure.

Entropy As a Measure of Political Uncertainty

Suppose instead of estimating respondent uncertainty by respondent and candidate characteristics we use the empirical distribution of placements across the survey. For example, if an issue was particularly complicated or it was one in which candidates generally avoided taking specific positions, then we would expect the distribution of candidate placements by respondents to be somewhat uniform across the range of answers reflecting aggregate voter uncertainty. Conversely, if the issue was one that was either very clear to the electorate or one where candidates tend to take unambiguous positions, then we would expect a highly concentrated distribution of respondent placements. What is required is a single value that can be substituted into (2) to replace $\text{Var}[C_{ijk}]$ based on this empirical measurement of uncertainty from the survey

As an example of the effects of uncertainty, Figure 1 shows how dramatically different respondents view candidate ideology depending on perceived uncertainty and party in the 1994 congressional election (data from the American National Election Study). The different empirical distributions demonstrate the value of measuring uncertainty. Not surprisingly, respondents indicating that they were not certain about ideology placement tended to place the candidates in the middle

Figure 1: Ideology Placement By Certainty, Respondent Counts by 7-Point Scale
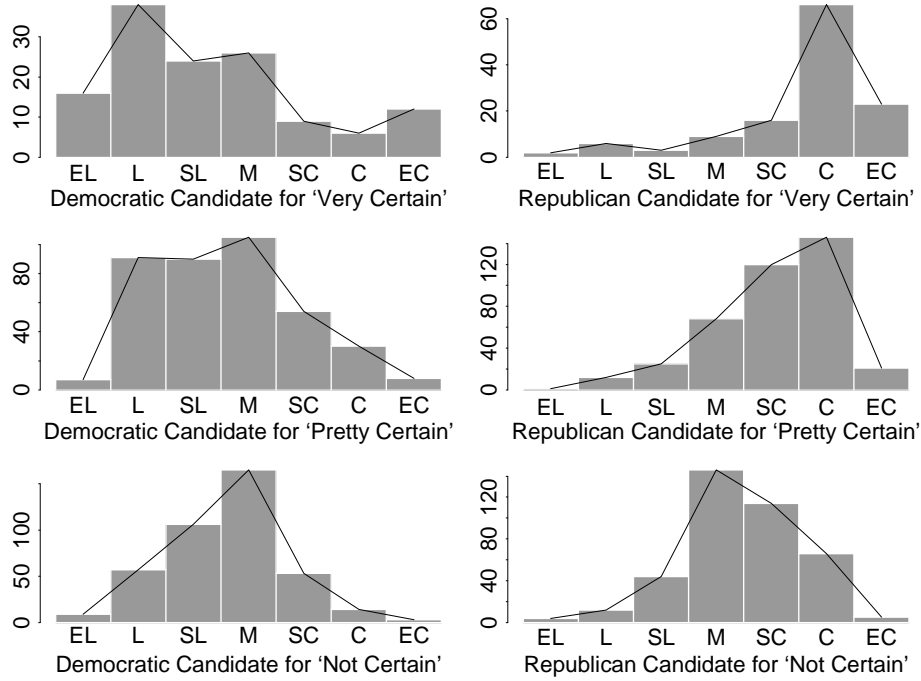
of the scale, clustered around the moderate category. Conversely respondents that were very sure about ideology placement were generally not reluctant to label candidates as more extreme. Figure 1 also indicates that there is less variance in the distribution of ideology placement for Republicans than Democrats.

The method proposed herein is to empirically assign the categorical probabilities (call these p_i for the probability of being in the ith category) for a given issue from the observed response structure in the dataset, and then calculate the entropy for that issue. So the measure is based on the issue, characteristics of the candidates, the wording of the question, and the aggregated responses. Thus the decision uncertainty is constant across respondents but not across candidates or issues.

The entropy approach makes absolutely no assumptions about the *distribution* of the variance of uncertainty (Shannon 1948; Jaynes 1968, 1982). Furthermore, empirical distributions that satisfy given linear constraints (side conditions), $\sum p_i = 1$ in our case, concentrate around the maximum entropy distribution which satisfies the same constraints (Van Campenhout and Cover 1981; Robert 1990). This finding, formalized as Jaynes' *Entropy Concentration Theorem* (1968), provides a motivation for treating the unknown distribution of candidate uncertainty as an entropy term because it builds directly on Shepsle's (1972) idea that observed candidate positions are realizations from an underlying probability distribution.

The entropy concept has been applied to a wide variety problems concerning uncertainty, yet there is no known application of entropy to measuring the lack of knowledge political actors have about each other's positions and intentions. Uncertainty about candidates and ballot issues is a fundamental problem for voters. Information is necessarily limited by transmission means (Shannon 1948), candidates and issues can be confusing at the source (Franklin 1991; Gerber 1996), and voters have limited time and attention and might search for shortcuts for even the most salient election issues (Carmines and Stimson 1980). The search for a measure of uncertainty is also disadvantaged in that uncertainty in individuals is a subjective and changing phenomenon (Alvarez and Franklin 1994).

The entropy approach to measuring voter uncertainty not only has few assumptions, but also applies to survey data where there are no specific uncertainty questions (Alvarez and Franklin 1994). Furthermore, there is no need to estimate a uniform variance threshold (Bartels 1988), and also no need to exploit some unique attribute of the data (Franklin 1991) if one exists.

Entropy in Information Theory

There are many definitions of information in various literatures but all of them have the same property of distinction from a message. If a message is the physical

manifestation of information transmission (language, character set, salutations, headers/trailers, body, etc.), then the *information* is the substantive content independent of transmission processes. For example, in baseball the third base coach transmits a message to the runner on second base through the use of hand-gestures. The information contained in the hand gestures is independent of the actual gesture to the base-runner. If you could condense a set of messages down to a minimal finite length, then the information content would be the number of digits required by these alternate sets of messages. In this case information and message would be identical. Information and message are never exactly the same. For example, the DNA strand required to uniquely determine the biological characteristics of a human being contains far more codes than is minimally sufficient to transmit the required genetic information. In information theory entropy is a measure of the average amount of information required to describe the distribution of some random variable of interest (Cover and Thomas 1991). More generally, information can be thought of as a measure of reduction in uncertainty given a specific message (Shannon 1948; Ayres 1994, p.44).

Suppose we wanted to identify a particular voter by serial information on this person's characteristics. We are allowed to ask a consecutive set of yes/no questions (i.e. like the common guessing game). As we get answers to our series of questions we gradually converge (hopefully, depending on our skill) on the desired voter. Our first question is: does the voter reside in California. Since about 13% of voters in the United States reside in California, a yes answer gives us different information than a no answer. Restated, a yes answer reduces our uncertainty more than a no answer because a yes answer eliminates 87% of the choices whereas a no answer eliminates 13%. If P_i is the probability of the ith event (residing in California), then the improvement in information is defined as:

$$I_{P_i} = \log_2 \left[\frac{1}{P_i} \right] = -\log_2 P_i. \tag{3}$$

The probability is placed in the denominator of (3) because the smaller the probability, the greater the investigative information supplied by a yes answer. The log function is required to obtain some desired properties (discussed below), and is justified by limit theorems (Bevensee 1993; Jaynes 1982; Van Campenhout and Cover 1981). The log is base-2 since there are only two possible answers to our question (yes and no) making the units of information bits. In this example $H_i = -\log_2(0.13) = 2.943416$ bits, whereas if we had asked: does the voter live in Arkansas then an affirmative reply would have increased our information by: $H_i = -\log_2(0.02) = 5.643856$ bits, or about twice as much. However, there is a much smaller probability that we would have gotten an affirmative reply had the question been asked about Arkansas. What Slater (1938) found, and Shannon (1948) later refined was the idea that the "value" of the question was the information returned by a positive response times the probability of a positive response. So the value of the ith binary-response question is just:

$$H_i = f_i \log_2 \left[\frac{1}{f_i} \right] = -f_i \log_2 f_i. \tag{4}$$

And the value of a series of n of these questions is:

$$\sum_{i=1}^{n} H_i = k \sum_{i=1}^{n} f_i \log_2 \left[\frac{1}{f_i} \right] = -k \sum_{i=1}^{n} f_i \log_2 f_i \tag{5}$$

where f_i is the frequency distribution of the ith yes answer and k is an arbitrary scaling factor that determines choice of units. The arbitrary scaling factor makes the choice of base in the logarithm unimportant since we can change this base by manipulating the constant.[1] The function (5) was actually first introduced by Slater (1938), although similar forms were referenced by Boltzmann (1877) where

---

[1]For instance if the entropy form were expressed in terms of the natural log, but $\log_2$ was more appropriate for the application (such as above), then setting $k = \frac{1}{\ln 2}$ converts the entropy form to base 2.

f_i is the probability that a particle system is in microstate i.

We can see that the total improvement in information is the additive value of the series of individual information improvements. So in our simple example we might ask a series of questions narrowing down on the individual of interest. Is the voter in California? Is the voter registered as a Democrat? Does the voter reside in an urban area? Is the voter female?. The total information supplied by this vector of yes/no responses is the total information improvement in units of bits since the response-space is binary. Its important to remember that the information obtained is defined only with regard to a well-defined question having finite, enumerated responses

The link between information and entropy is often confusing due to differing terminology and usage in various literatures. In the thermodynamic sense, information and entropy are complementary: "gain in entropy means loss in information - nothing more" (Lewis 1930). So as uncertainty about the system increases, information about the configuration of the microstate of a system decreases. In terms of information theory, the distinction was less clear until Shannon (1948) presented the first unambiguous picture. To Shannon the function (5) represented the *expected* uncertainty. In a given communication, possible messages are indexed $m_1, m_2, \ldots, m_k$, and each have an associated probability $p_1, p_2, \ldots, p_k$. These probabilities sum to one, so the scaling factor is simply 1. Thus the Shannon entropy function is the negative expected value of the natural log of the probability mass function, with a scaling factor c:

$$H = -c \sum p_i \ln(p_i). \tag{6}$$

Shannon defined the information in a message as the difference between the entropy before the message and the entropy after the message. If there is no information before the message, then a uniform prior distribution[2] is assigned to the p_i

---

[2]The uniform prior distribution as applied provides the greatest entropy since no single event is

and entropy is at its maximum[3] In this case any result increases our information. Yet, if there is certainty about the result, then a degenerate distribution describes the m_i, and the message does not change our information level.[4]. So according to Shannon, the message produces a new assessment about the state of the world and this in turn leads to the assignment of new probabilities (p_i) and a subsequent updating of the entropy value.

The simplicity of the Shannon entropy function belies its theoretical and practical importance. Shannon (1948, Appendix 2) showed that (6) is the only function that satisfies the following three desirable properties:

1. H is continuous in $(p_1, p_2, \ldots, p_n)$.

2. If the p_i are uniformly distributed, then H is at its maximum and is monotonically increasing with n.

3. If a set of alternatives can be reformulated as multiple, consecutive sets of alternatives, then the first H should equal the weighted sum of the consecutive H values: $H(p_1, p_2, p_3) = H(p_1, 1 - p_1) + (1 - p_1)H(p_2, p_3)$.

These properties mean that the Shannon entropy function is well-behaved with regard to relative information comparisons.[5]

---

more likely to occur. Thus the uniform distribution of events provides the minimum information possible with which to decode the message. This application of the uniform distribution does not imply that this is a "no information" assumption since equally likely outcomes is certainly a type of information. A great deal of controversy and discussion has focused around the erroneous treatment of the uniform distribution as a zero-based information source (Dale 1991; Stigler 1986).

[3]$H = -\sum \frac{1}{n} \ln(\frac{1}{n}) = \ln(n)$, so entropy increases logarithmically with the number of equally likely alternatives.

[4]$H = -\sum_{i=1}^{n-1}(0) - \log(1) = 0$

[5]Only the discrete form of the entropy formulation has been discussed. Since entropy is also a measure of uncertainty in probability distributions, the following entropy formula also satisfies all of the properties enumerated above: $H = \int_{-\infty}^{+\infty} f(x)\ln(f(x))dx$. This form is often referred to as the *differential entropy* of a random variable X with a known probability density function $f(x)$. There is one important distinction: the discrete case measures the absolute uncertainty of a random variable given a set of probabilities, whereas the continuous case this

There has been considerable debate about the nature of information with regard to entropy (Ayres 1994; Jaynes 1957; Ruelle 1991; Tribus 1961, 1979). A major question arises as to *whose* entropy is it? The sender? The receiver? The transmission equipment? These are often philosophical debates rather than practical considerations as Shannon's focus was clearly on the receiver as the unit of entropy measure. Jaynes (1982) however considers the communication channel and its limitations as the determinants of entropy. Some, like Aczél and Daróczy (1975), view entropy as a descriptor of a stochastic event, and the traditional thermodynamic definition of entropy is as a measure of uncertainty in microstates. For the purposes of this work, the Shannon formulation (6) is used.

An Entropy Model of Vote Choice

For each issue in which there exists in the survey a candidate assessment and a self-placement, it is possible to provide an entropy measure. The entropy term for a given issue is created from the empirically observed counts in each response category and then normalized into probabilities. Returning to the 1994 American National Election Study data, the Democratic candidates' respondent placement distribution on support for Clinton is the normalized vector:

$$[0.077, 0.014, 0.000, 0.899, 0.000, 0.007, 0.003].$$

These are just the normalized probabilities that a respondent selected the corresponding categorization for the Democratic candidates perceived support of Clinton (from almost always supports Clinton to almost never supports Clinton). This produces an entropy value of $H_{\text{Democrat,Support.Clinton}} = 0.404$ by applying (6). In contrast, a uniform distribution produces the maximum entropy

---

uncertainty is measured relative to the chosen coordinate system (Ryu 1993). So transformations of the coordinate system require transformation in the probability distribution function with the associated Jacobian. In addition, the continuous case can have infinite entropy unless additional, "side" conditions are imposed (see Jaynes 1968).

13

value for a seven category discrete random variable: $H = 1.946$. Note that like the variance, the entropy measure of uncertainty is "direction-less" since $[0.5, 0, 0.5, 0, 0, 0]$ and $[0, 0, 0, 0.5, 0, 0.5]$ return the same value. Unlike the variance, the entropy measure retains uncertainty in the placement of the scale, $[0.5, 0, 0, 0, 0, 0.5]$, $[0, 0, 0.5, 0, 0.5, 0, 0]$ return the same value. In both of these latter cases, respondents should be equally uncertain as to where the candidates stand; just because the second candidate has closer modal values does not imply less uncertainty in the context of a 7-point scale.

For each of the K issue terms and candidates an entropy function, H_{jk} is created to indicate the corresponding decrease in utility from uncertainty. We can directly substitute the entropy term for the variance term in (2) to give the expected utility of candidate j to voter i as:

$$E[U_{ij}] = \sum_{k=1}^{K} \left[-(\mathrm{E}[C_{ijk}] - R_{ik})^2 - H_{jk} \right] + O_{ij} \tag{7}$$

where the H_{jk} is not indexed by i since it comes from the entire dataset. Entropy and issue distance are additive here, so they must be standardized on the same scale for each issue. Furthermore, since every variable was standardized on the 7-point scale, relative comparisons across covariates in the model results also make sense.

The highlighted difference between (7) and Bartels' formulation in (2) is that the entropy measure reflects the full sample attitude about uncertainty varying by issue and candidate. The uncertainty component is provided by a measure of all respondents' summed assessments through the candidate/issue entropy term in (7). Unless one is willing to assert that uncertainty is systematically related to observed respondent characteristics, then there is literally nothing in the data to obtain information about respondents' level of candidate position uncertainty.

The entropy measure takes advantage of questions inserted into the ANES

14

1995 Pilot Study directed at assessing how certain a respondent is about the candidates positions on a subset of issues. Respondents could indicate whether they were "very certain," "pretty certain," or "not very certain" after answering a particular substantive question. This allows us to build an aggregate measure of uncertainty by candidate and by issue that captures the effects in the survey sample without assuming direct relationships to covariates. While an individual's level of uncertainty guides their vote choices, the entropy approach gives an aggregated picture of the vagueness or specificity from issues and candidates that feeds into that. Thus it is a valuable component for the statistical model.

The form of (7) implies that a particular issue can reduce respondent utility either through squared issue distance from the candidate *or* from increasing the level of uncertainty expressed through the entropy term. The j and k subscript on the entropy term specify the corresponding candidate and issue distribution from which the entropy term is calculated. So this term is constant across respondents but not candidates and issues.

After calculating the complete utility function for two competing candidates, a measure of quality of the model and assumptions is the degree to which prediction agrees with observed vote choice. Specifically, we would expect for a particular voter i, if $\mathrm{E}[U_{i1}] > \mathrm{E}[U_{i2}]$, then voter i selects candidate 1. The form of (7) incorporates the *threshold existence property*: issue based utility preference for one candidate over another is subject to mitigation by high levels of uncertainty (Brady and Ansolabehere 1989), where the uncertainty component is measured here by entropy rather than the unobtainable statistical variance threshold.

A key feature of this structural specification is that the entropy and issue distance are *normalized* to be on the same scale and thus allowed to compete. The question we are asking, therefore, is whether issue distance is sufficient to overcome uncertainty, or whether uncertainty can be substantial enough to overwhelm issue distance. The empirical test of these questions involves building

15

a statistical model in which each candidate is assessed across issues where the respondent places this candidate *and* themselves, and where we can obtain the entropy distribution from aggregate uncertainty controlling for demographics and other personal factors. Distance matters, therefore, because of both issue placements and candidate entropy. So for the issues assessed in this way (five actually, given the data at hand), there will be a Democratic term and a Republican term, which are actually directly comparable. The advantage to this approach is that it not only allows direct issue-by-candidate evaluation where the estimated coefficient provides an importance weight in the decision making process, but also that it avoids the problems described above in obtaining variance thresholds.

In both (2) and (7) the issues are equally weighted with regard to their contribution to total utility. This is not actually a requirement. If there was some reason for assuming that certain issues were more important or more topical than others, then deliberately specified salience weights could easily be assigned to each of the K issue terms reflecting their a priori relative importance. This is not done for the five issue distance variables considered below, because there is no substantive justification for claiming that any one of the policy areas supersedes the others in terms of relevance to the election and we will estimate their relative importance empirically (the coefficients can be thought of as posterior weights in our context). Also, these weights could be estimated in the more general context of a Bayesian hierarchical model where their values are stipulated as conditional on additional criteria such as individual voter characteristics. This approach is not particularly difficult if one is willing to adapt the tenets of Bayesian inference which have considerable power and elegance (Gill 2002).

Heteroscedasticity and the Likelihood Function

Attitudes and assessments about political figures are likely to vary across respondents such that the residuals from a simple generalized linear model are heteroscedastic. A graphical diagnostic test of the model described above indicates

that this is the case here. It is not surprising to find that respondents perceive choices and candidate issue differences in this particular election with non-uniform variation.

Because we are interested in an expected utility based choice between two candidates, a dichotomous model specification is appropriate. Heteroscedasticity is particularly troublesome in qualitative choice models as the resulting maximum likelihood estimates of the coefficients are often inconsistent (Amemiya 1981; Davidson and MacKinnon 1984) and the asymptotic covariance matrix (generally computed from the inverse of the negative Hessian at the maximum likelihood coefficient values) is biased. Yatchew and Griliches (1985) also point out that the biases are particularly problematic when an explanatory variable is correlated with the heteroscedasticity (p.138).

An effective way of compensating for heteroscedasticity in probit models is to identify explanatory variables that introduce widely varying response patterns and specifically associate them in the model with the dispersion term. This method, based on Harvey (1976), is illustrated in Greene (2003, p.680-1), and was recently employed by Alvarez and Brehm (1995, 1997) to address response heteroscedasticity from issue ambivalence. The functional form is changed from:

$$P(Y = 1) = \Phi(\mathbf{X}\boldsymbol{\beta}) \qquad \text{to:} \qquad P(Y = 1) = \Phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{e^{\mathbf{Z}\boldsymbol{\alpha}}}\right) \qquad (8)$$

where Φ is the standard notation for the cumulative standard normal distribution, and the exponent function in the denominator is just a convenience that prevents division by zero. This approach distinguishes between the standard treatment of the estimated coefficients $\boldsymbol{\beta}$ with corresponding matrix of explanatory observations $\mathbf{X}$, and the set of dispersion determining estimated coefficients $\boldsymbol{\alpha}$ with a corresponding matrix of explanatory observations $\mathbf{Z}$. By this means the dispersion term is reparameterized to be a function of a set of coefficients and observed values that are suspected of causing the differences in error standard deviations:

$\sigma_i^2 = e^{\mathbf{Z}_i\boldsymbol{\alpha}}$. The log likelihood of the probit model is now:

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}|\mathbf{y}, \mathbf{X}, \mathbf{Z}) = \sum_{i=1}^{n} \left[y_i \log\left(\Phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{e^{\mathbf{Z}\boldsymbol{\alpha}}}\right)\right) + (1 - y_i) \log\left(1 - \Phi\left(\frac{\mathbf{X}\boldsymbol{\beta}}{e^{\mathbf{Z}\boldsymbol{\alpha}}}\right)\right) \right] \quad (9)$$

where the inclusion of a separate parameterized dispersion component in (9) addresses the heteroscedasticity problem. A formal test for heteroscedasticity is a likelihood ratio test using the value of the log likelihood in the restricted model (the restriction is that $\boldsymbol{\alpha} = \mathbf{0}$) and the unrestricted (9) evaluated at the maximum likelihood estimates of the parameters: $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. The test is simply: $LR = -2\left[\ell_{\text{restricted}} - \ell_{\text{unrestricted}}\right] = 2\left[\ell(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\boldsymbol{\alpha}) - \ell(\mathbf{X}\boldsymbol{\beta})\right]$, where LR is (as usual) distributed χ^2 but with degrees of freedom equal to the number of columns of the $\mathbf{Z}$ matrix (the number of dispersion estimation parameters in the Harvey specification).

Testing the Entropy Model of Uncertainty

The 1994 congressional election represents an interesting and perhaps unique opportunity to test the effect of uncertainty on vote choice. Republican House candidates signed the *Contract with America* on September 9, 1994 with the goal of capturing the majority in Congress by specifically identifying their policy positions through a set of ten promised House floor votes. It is reasonable to predict that a clearly defined, written agenda that is widely reported in the media increases voter certainty about the Republican candidates' positions. We test this claim by applying the entropy model described above using survey data from the 1994 American National Election Study (data collected between November 9, 1994 and January 9, 1995).

The data include placements for 1,795 respondents on several currently salient issues as well as standard demographic information. Fortunately, for five of these issue questions the respondent is asked not only to place themselves, but also to place the Republican and Democratic House candidates running in their district

18

(both only if this is a contested race). The five issues are: support for Clinton, concern about crime, the government's role in providing economic opportunities to citizens, appropriate level of government spending, and the appropriate level of government involvement in the healthcare system. These variables are particularly useful since they are policy areas addressed by the 103rd Congress and the respondent places themself and the two candidates for the House. Each of these fifteen variables (five times each for the respondent, the Republican, and the Democrat) are rescaled to the seven-point format, and those responding "Don't Know" for the issue variables are placed at the midpoint of the scale in order to minimize their squared error distance and subsequent effect on the model (details provided in the Data Appendix). The scales are standardized such that higher values are those traditionally associated with Republican positions. For instance, the question that asks about increasing or decreasing government spending for the provision of domestic services was recoded to make "7" the response indicating government should provide far fewer services.

Four general demographic variables are included in the model: *Education* (in years), *Gender*, *Race* (white or non-white), and *Age* (in years). The model also includes two variables which are intended to capture the respondents' levels of interest and information. The data include a question which asks the respondents to place themselves on a three-point scale in terms of their attention to recent campaigns. Respondent information is measured by mean-averaging respondents' number of days reading the newspaper per week, number of days watching television per week, and number of days listening to news on the radio per week.

Three of the explanatory variables are specified as belonging to the $\mathbf{Z}$ matrix in $\left(\frac{\mathbf{X}\boldsymbol{\beta}}{e^{\mathbf{Z}\boldsymbol{\alpha}}}\right)$: *Party Identification Scale*, *Political Interest Scale*, and *Education*. It hypothesized that self-reported partisanship tends to vary the effects of issue-distance since respondents will incorporate party loyalties differently in assessing distances. For different reasons, levels of interest in politics and levels of education

should alter the quantity and quality of political information that an individual obtains. All of these variables are subsequently mean-centered in the analysis so that $\exp(\mathbf{Z}\boldsymbol{\alpha}) = 1$ at the mean. This reduces the sensitivity of the $\boldsymbol{\beta}$ parameters to scale standardization.

Multiple imputation is used to replace the explanatory variable missing values resulting from refusals and survey errors (Little and Rubin 1983; Rubin 1987). Cases where the House race is for an open seat were dropped since the survey therefore does not provide data on support for Clinton and vote on the Crime bill for non-incumbent candidates. These considerations and the treatment of the outcome variable (described below) reduce the utilized sample size to 1,335.[6]

The outcome variable is candidate choice in the 1994 House election by party, where Democrats are coded 0, and Republicans are coded 1. Respondents who did not claim to have voted were excluded from the analysis, although very similar results were obtained by replacing their missing votes with their stated vote *preferences* and including these respondents in the sample. So there are exactly two utility functions from (7) for each respondent: U_{iD} for the ith respondent's utility from the Democratic candidate, and U_{iR} for the ith respondent's utility from the Republican candidate. The inferential value of this qualitative choice model is assessed by the frequency with which the respondent selects the candidate with the greater utility function. Table 1 summarizes the results from the heteroscedastic probit regression model, (9), on these data.

Findings

The heteroscedasticity likelihood ratio test indicates that we are justified in developing the Harvey heteroscedastic probit model for these data over a standard specification: a test statistic of 101 $(\sim \chi^2_{df=3})$ is clearly sufficient evidence to reject the hypothesis that $\boldsymbol{\alpha} = \mathbf{0}$. Furthermore, the residual deviance, 1282.6, is not in

---

[6]All data, code, and imputations for replication purposes are archived at the website: .

20

21

Table 1: Parameter Estimates for Entropy Model of Vote Choice 1994 ANES

	Parameter Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound
Choice Parameters ($\mathbf{X}\boldsymbol{\beta}$)				
Intercept	-0.4777	0.2096	-0.8885	-0.0669
Democratic Support for Clinton	0.0115	0.0063	-0.0008	0.0238
Republican Support for Clinton	0.0112	0.0074	-0.0034	0.0257
Democratic Crime Concern	0.0102	0.0065	-0.0025	0.0230
Republican Crime Concern	-0.0206	0.0064	-0.0331	-0.0081
Democratic Government Help for Disadvantaged	0.0285	0.0080	0.0128	0.0442
Republican Government Help for Disadvantaged	-0.0155	0.0085	-0.0322	0.0012
Democratic Government Spending	0.0996	0.0185	0.0632	0.1359
Republican Government Spending	-0.0904	0.0141	-0.1180	-0.0628
Democratic Federal Healthcare	0.0150	0.0054	0.0045	0.0256
Republican Federal Healthcare	-0.0145	0.0060	-0.0262	-0.0028
Media Exposure	-0.0202	0.0068	-0.0336	-0.0068
Gender	-0.3332	0.0769	-0.4838	-0.1826
Race	-0.2615	0.1010	-0.4595	-0.0635
Age	0.0019	0.0021	-0.0023	0.0061
Dispersion Parameters ($e^{\mathbf{Z}\boldsymbol{\alpha}}$)				
Party Identification Scale	0.1951	0.0168	0.1623	0.2280
Political Interest Scale	-0.2307	0.0332	-0.2957	-0.1657
Education	0.0495	0.0240	0.0025	0.0966

Likelihood Ratio Test: $LR = 82.35$, $p < 0.001$ for $\chi^2_{df=3}$
Correctly Classified: 946/1335 (using the mean-criteria)
Residual Deviance = 1282.6

Outcome Variable:
Democrats =0, Republicans = 1

the tail of a χ^2 distribution for 1317 degrees of freedom, justifying the model fit.

The five pairs of issue terms (the first ten terms after the constant in Table 1 are interpreted as follows. Take, for instance, *Support for Clinton* as the $k = 1$ issue and $j = 1$ for the Democratic House candidate, to be considered by the ith respondent. The first column of the $\mathbf{X}$ matrix (save for the column of ones) is constructed by: $-(C_{i11} - R_{i1})^2 - H_{11}$. So this column of data is constructed from the negative square of the distance between self-placement and candidate placement, minus the entropy uncertainty for this issue and partisan candidate. Thus the explanatory variable's value is the degree to which it explains variation of the outcome variable (on the probit scale), *and* the degree to which it can overcome corresponding entropy uncertainty.

The individual coefficient results in Table 1 suggest that the model does reflect vote choice under uncertainty in a reasonable manner. Nine of the thirteen choice coefficients in the numerator of the (8), have 95% confidence intervals bounded away from zero. Of the partisan issue distance pairs, all are signed in the anticipated direction, negative for Republicans and positive for Democrats, except for *Support for Clinton* (both of which are not reliable anyway). Recall that a positive coefficient pushes respondents away from Democrats and a negative coefficient pushes respondents away from Republicans since these are based on squared issue distance plus entropy uncertainty.

Government spending policy was an important issue in the 1994 campaign, and the model produces results exactly as we would predict. Spending continues to be a potent campaign topic because it allows House members to easily run against Congress as an institution (Fenno 1977). The finding that the Democratic and Republican coefficients are of equal magnitude is interesting as it suggests that spending distance, in the presence of uncertainty, affects both parties in similar fashion.

Part of the Republican agenda enshrined in the *Contract with America* was

the idea of reigning-in the Clintons' healthcare initiative. We find nearly identical effects of issue distance for Democrats and Republicans, indicating again that issue increasing distance drives voters away in the same way.

The three choice covariates in the denominator of the Harvey specification produce reliable coefficients with expected effects: females and non-whites are more likely to vote Democratic, and increased media consumption leads to more Democratic voting. While these are control variables in our context, such reliability and predictability is certainly reassuring about the overall quality of the model.

Unfortunately *Age* is not statistically reliable here. This may be because we are controlling for enough of the other covariates related to the respondents' age, such as education, political interest, and healthcare, that age lacks differentiated explanatory power.

Conclusion

It appears that uncertainty matters. This model demonstrates that in the aggregate, including a measure of uncertainty affects the importance of individual issue-distances. Uncertainty matters because it does or does not mitigate issue distance on a per-issue basis in the classical sense from models such as those given by Enelow and Hinich (1981), Shepsle (1972) and Bartels (1988). The implication is that politicians gain by clarifying policy positions provided that those positions are not too far out of line with those widely supported by the electorate. Or, at a minimum, there is evidence that politicians need to trade-off the cost of uncertainty with the cost of alienating specific groups of voters.

The three dispersion covariates in the denominator of the Harvey specification also produce reliable coefficients. As expected, the *Party Identification* coefficient is large, positive, and has a 95% confidence interval which does not cover zero. There is a vast literature to support this finding even though the effects appear to

23

be waning (Nie, Verba and Petrocik 1976; Schulman and Pomper 1976; Page and Jones 1979; Wattenberg 1990). It appears that increased levels of political interest decrease model variability and subsequently increase the reliability of the choice parameters since the coefficient for *Political Interest Scale* is negative and reliable. This makes intuitive sense since those with greater interest are more likely to seek candidate and issue information than less-interested individuals. The finding for *Education* is curious since the corresponding coefficient is positive and reliable, and therefore gives the opposite effect.

Two other issues give only partially satisfactory results. In the partisan co-efficient pairs for *Government Help for Disadvantage* and *Crime Concern*, one coefficient is reliable at the 95% level and the other is not possibly because the contribution from the entropy uncertainty term exceeds that of the spatial distance term, or for purely statistical reasons (lack of sample information, given this model).

Crime has been a bread-and-butter Republican campaign issue for decades, and the model indicates that increased distance, factoring in uncertainty, reduces the probability that the respondent will vote for the Republican candidate. Notice also that while the standard error for the Democratic candidate is identical (subject to rounding), the Republican coefficient is about twice that of the other, which for this reason is not statistically reliable. Taken together, this suggests that Republicans are almost certainly punished for failing to take a position close to the electorate while there is no evidence that Democrats must do so (although the coefficient is signed in the direction we would expect).

Interestingly, the bread-and-butter *Democratic* campaign issue of help for societies shows exactly the pattern as seen with crime, except that the roles are reversed. The coefficient for the Democratic candidate is reliable while that for the Republican is not (and once again the standard errors are comparable). In addition, just like crime, the reliable coefficient is twice that of the other suggest-

24

ing that Democrats are punished for deviating sharply from positions close to the electorate but there is no evidence that Republicans suffer the same fate.

The model provides interesting and somewhat perplexing results with regard to the *Support for Clinton* variable. Both the Democratic and the Republican variables have 95% confidence intervals which bound zero so we cannot place much faith in these findings, but the Republican variable is signed oppositely than expected: greater distance between the Republican candidate and respondent provides greater probability of voting for that candidate. We would normally expect that as the distance increases, a respondent is increasingly dissatisfied with the Republican candidate and more likely to vote for the Democratic candidate. It is possible, but not able to be shown here, that there is something special happening with this issue, and the lack of clarity supports Clinton's status as somewhat of an enigma. Clinton often embraced traditional Republican positions as a "New Democrat" (reinventing government, a focus on the economy, protection of programs for the elderly, reevaluation of welfare policies), but he also alienated many Republicans with his stance on gays in the military and other issues.

Entropy as a measure of uncertainty in political communication indicates a lack of substance in transmissions. Entropy captures the cost of uncertainty in this calculus by measuring this uncertainty through analysis of discrete issue placements. The entropy measure can be superior to other tools because it calculates uncertainty with reasonable assumptions: diffusion is equivalent to uncertainty, information is the substantive component independent of the transmission process, and the receiver (i.e. voter) is the judge of information quality and quantity. Conversely, measures of information uncertainty based on variances require indirect estimation of unobservable parameters and therefore further specification.

Rather than find evidence that the *Contract with America* made a substantive difference in the electoral fortunes of the Republican candidates, the conclusion from this research is that combined issue distance and uncertainty disproportion-

25

ately affect Democrats, thus diminishing the importance of the Republican campaign pledge. The asymmetry of the signed coefficients slightly favors Republicans over Democrats. It is possible that forty years of House control created a high level of *perceived* certainty about the policy positions of Democratic candidates and for them issue distance was more important than vagueness or clarity.

The *Contract with America* was a Republican effort to stake out and explicate a set of specific policy positions. The political sophistication of the Republican leaders in the House during this time remains open to question. If the *Contract with America* was developed to increase the clarity of Republican candidates' positions nationwide, then it would imply a high level of strategic foresight with regard to analyzing voter preferences and uncertainty. However, there are other reasons that the leadership developed an explicit agenda and required members to publicly sign the document. These include control of the agenda by the leadership, control of the freshmen members, and a strategy for opposing policies advocated by the president. It is entirely possible that the Republican leadership was not aware of the possible advantage available by reducing uncertainty, but they simply benefited by luck.

Data Appendix

This section summarizes the data format and coding decisions applied to data from the 1994 American National Election Study (study No.6507 conducted between November 9, 1994 and January 9, 1995). Categorical sums given for each outcome exclusive of missing values. For the five pairs of issue variables below ($k = 1 : K$), three scales are created: one for the respondent's self-placement (R_{ik}), and one each for the respondent's placement of their corresponding Democratic candidate and Republican candidate ($C_{ijk}, j = 1, 2$). The aggregated candidate placements are used to create the corresponding entropy value (H_{jk}). Demographic covariates ($O_{ij}, j = 3 : 9$) are left in their original measurement, with the adjustments described below.

1. *Support for Clinton.* The respondent measure is specified as a seven-point scales by combining the approve/disapprove question (`VAR 201`) with the strength of approval/disapproval question (`VAR 202`). The seven-point *Democratic Candidate Support for Clinton* variable and *Republican Candidate Support for Clinton* variables were created by combining House incumbent supports Clinton Half

26

the time (`VAR 647`) with House incumbent supports Clinton almost always (`VAR 645`) and House incumbent supports Clinton almost never (`VAR 646`) then cross-referencing by incumbent's party (`VAR 17`).

2. *Concern About Crime.* These measures for the respondent and the two candidates are created by rescaling the voted-for or supported the Crime Bill question (`VAR 1040` for each respondent, `VAR 648` and `VAR 649` for the incumbent, cross referenced again by incumbent's party), where the respondents' interior points are derived from the question asking about appropriate level of spending on programs dealing with crime (`VAR 825`). This procedure gives the desired seven-point scale.

3. *Government Help for Disadvantaged.* The three measures come directly from self-assessment (`VAR 930`) as well as placement of the Democratic (`VAR 932`) and Republican (`VAR 933`) candidates on a seven-point scale where 1 indicates "the government should see to a job and good standard of living," and 7 indicates "government should let each person get ahead on their own." No rescaling is necessary.

4. *Government Spending.* The three measures of government spending on social programs come directly from self-assessment (`VAR 940`) as well as placement of the Democratic (`VAR 942`) and Republican (`VAR 943`) candidates on a seven-point scale where 1 indicates "the government should many fewer services," and 7 indicates "the government should many more services." No rescaling is necessary, but the scale direction is reversed to put the more typically Republican position on the high end of the scale.

5. *Federal Healthcare.* The three healthcare measures also come from self-assessment (`VAR 950`) as well as placement of the Democratic (`VAR 952`) and Republican (`VAR 953`) candidates on a seven-point scale where 1 indicates that the provision of healthcare nationally should be completely through a "government insurance plan" and 7 indicates it should be completely through a "private insurance plan" No rescaling is necessary.

6. *Party Identification Scale.* This is given as a seven-point summary scale (`VAR 655`) with the categories: (0) strong Democrat, (1) weak Democrat, (2) independent, leaning Democrat, (3) independent, (4) independent, leaning Republican, (5) weak Republican, (6) strong Republican.

7. *Political Interest Scale.* Political interest is measured directly on the points: (1) very interested, (2) somewhat interested, (3) not much interested (`VAR 124`).

8. *Media Exposure.* This variable is specified by averaging daily consumption across three media sources. The respondent reports the typical number of days per week that they: read the paper (`VAR 205`), watch TV news (`VAR 206`), and listen to radio news (`VAR 207`).

9. *Education.* The ANES summary measure (`VAR 1209`) is used in its provided form with seven categories: 8th grade or less, 9th-11th grade, high school or GED, more than 12th, associate degree, bachelors degree, and advanced degree.

10. *Gender.* Gender is switched in order from `VAR 1434` such that females are coded 0 and males are coded 1.

11. *Race.* Race is dichotomized according to *white* (0) and *not-white* (1) from `VAR 1435`.

12. *Age.* Actual reported age is used (`VAR 1203`), ranging from 18 to 91, with mean 46.5, and standard deviation 18.

13. *Vote Choice.* The outcome variable is dichotomized by collapsing the information from `VAR 613` such that voting for the Democrat is coded 0, and voting for the Republican is coded 1.

References

Abramson, Paul R. and John H. Aldrich. 1982. "The Decline of Electoral Participation in America." *American Political Science Review* 76: 502-21.

Aczél J. and Z. Daróczy. 1975. *On Measures of Information and Their Characterizations.* New York: Academic Press.

Aldrich, John H., Richard G. Niemi, George Rabinowitz, and David W. Rohde. 1982. "The Measurement of Public Opinion about Public Policy: A Report on Some New Issue Question Formats." *American Journal of Political Science* 26: 391-414.

Alvarez, R. Michael. 1998. *Issues and Information in Presidential Elections.* Second Edition. Ann Arbor: University of Michigan Press.

Alvarez, R. Michael and John Brehm. 1995. "American Ambivalence Towards Abortion Policy: Development of a Heteroscedastic Probit Model of Competing Values." *American Journal of Political Science* 39: 1055-82.

Alvarez, R. Michael and John Brehm. 1997. "Are Americans Ambivalent Towards Racial Policies?" *American Journal of Political Science* 41: 345-72.

Alvarez, R. Michael and Charles H. Franklin. 1994 "Uncertainty and Political Perceptions." *Journal of Politics* 56 (3): 671-88.

Amemiya, Takeshi. 1981. "Qualitative Response Models: A Survey." *Journal of Economic Literature* 19: 1483-536.

Ayres, David. 1994. *Information, Entropy, and Progress.* New York: American Institute of Physics Press.

Bartels, Larry. 1996. "Uninformed Votes: Information Effects in Presidential Elections." *American Journal of Political Science* 40: 194-230.

Bartels, Larry. 1988. "Issue Voting Under Uncertainty: An Empirical Test." *American Journal of Political Science* 30: 709-28.

Bevensee, Robert M. 1993. *Maximum Entropy Solutions to Scientific Problems.* Englewood Cliffs, NJ: Prentice Hall.

Boltzmann, L. 1877. "Uber die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Warmetheorie und der Wahrscheinlichkeitsrechnung, respective den Satzen uber das Wermegleichgewicht." *Wein Ber* 76: 373-95.

Brady, Henry and Stephen Ansolabehere. 1989. "The Nature of Utility Functions in Mass Publics." *American Political Science Review* 83: 143-63.

Burnham, Walter Dean. 1987. "The Turnout Problem." In *Elections American Style.* ed. A. James Reichley Washington, D.C.: Brookings Institution.

Carmines, Edward G. and James A. Stimson. 1980. "The Two Faces of Issue Voting." *American Political Science Review* 74: 78-91.

Coombs, Clyde H. and Lolagene Coombs. 1976. "'Don't Know': Item Ambiguity or Respondent Uncertainty."' *Public Opinion Quarterly* 40: 497-514.

Coughlin, Peter J. 1990. "Candidate Uncertainty and Electoral Equilibria." In *Advances in the Spatial Theory of Voting*, eds. James M. Enelow & Melvin J. Hinich. Cambridge: Cambridge University Press.

Cover, Thomas M., Joy A. Thomas. 1991. *Elements of Information Theory.* New York: John Wiley and Sons.

Dale, Andrew I. 1991. *A History of Inverse Probability: From Thomas Bayes to Karl Pearson.* New York: Spring-Verlag.

Davidson, Russell and James G. MacKinnon. 1984. "Convenient Specification Tests for Logit and Probit Models." *Journal of Econometrics.* 25: 241-62.

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

Enelow, James M. and Melvin J. Hinich. 1990. *Advances in the Spatial Theory of Voting.* Cambridge: Cambridge University Press.

Enelow, James M. and Melvin J. Hinich. 1984a. *Spatial Analysis of Elections.* New York: Cambridge University Press.

Enelow, James M. and Melvin J. Hinich. 1984b. "Ideology, Issues, and the Spatial Theory of Elections." *The American Political Science Review* 76: 493-501.

Enelow, James M. and Melvin J. Hinich. 1981. "A New Approach to Voter Uncertainty in the Downsian Spatial Model." *American Journal of Political Science* 25: 483-93.

Feddersen, Timothy J. and Wolfgang Pesendorfer. 1999. "Abstention in Elections with Asymmetric Information and Diverse Preferences." *American Political Science Review* 93, 381-98.

Feddersen, Timothy J. and Wolfgang Pesendorfer. 1997. "Voting Behavior and Information Aggregation in Elections With Private Information." *Econometrica* 65, 1029-58.

Feddersen, Timothy J. and Wolfgang Pesendorfer. 1996. "The Swing Voter's Curse." *American Economic Review* 86, 408-24.

Fenno, Richard. 1977. "U.S. House Members in Their Constituencies: An Exploration." *American Political Science Review* 61, 883-917.

Ferejohn John A. and Morris Fiorina. 1974. "The Paradox of Not Voting: A Decision Theoretic Analysis." *American Political Science Review* 68: 525-36.

Franklin, Charles. 1991. "Eschewing Obfuscation? Campaigns and the Perceptions of U.S. Senate Incumbents." *American Political Science Review* 85: 1193-214.

Gafke, Roger and David Leuthold. 1979. "The Effect on Voters of Misleading, Confusing, and Difficult Ballot Titles." *Public Opinion Quarterly* 43: 394-401.

Gerber, Elisabeth. 1996. "Legislatures, Initiatives, and Representation: The Effects of State Legislative Institutions on Policy." *Political Research Quarterly* 49: 263-86.

Gill, Jeff. 2002. *Bayesian Methods: A Social and Behavioral Sciences Approach*; Chapman & Hall: New York.

Glazer, Amihai. 1990. "The Strategy of Candidate Ambiguity." *American Political Science Review* 84: 237-41.

Grafstein, Robert. 1991. "An Evidential Decision Theory of Turnout." *American Journal of Political Science* 35: 989-1010.

Greene, William H. 2003. *Econometric Analysis.* Fifth Edition. Saddle River, NJ: Prentice Hall.

Harvey, A.C. 1976. "Estimating Regression Models with Multiplicative Heteroscedasticity." *Econometrica* 44: 461-5.

Hinich, Melvin J. 1977. "Equilibrium in Spatial Voting: The Median Voter Is An Artifact." *Journal of Economic Theory* 16: 280-19.

Jaynes, Edwin T. 1957. "Information Theory and Statistical Mechanics." *Physics Review* 106: 620-30.

Jaynes, Edwin T. 1968 "Prior Probabilities." *IEEE Transmissions on Systems Science and Cybernetics* SSC(4): 227-41.

Jaynes, Edwin T. 1982. "On the Rationale of Maximum-Entropy Methods." *Proceedings of the IEEE* 70(9): 939-52.

Lewis, Gilbert N. 1930. "The Symmetry of Time in Physics ." *Science* LXXI: 569-77.

Little, Roderick J. A., and Donald B. Rubin. 1983. "On Jointly Estimating Parameters and Missing Data by Maximizing the Complete-Data Likelihood." *The American Statistician* 37, 218-20.

Nie, Norman H., Sidney Verba and John R. Petrocik. 1976. *The Changing American Voter.* Cambridge, MA: Harvard University Press.

Nimmo, Dan. 1970. *The Political Persuaders.* Englewood Cliffs, NJ: Prentice-Hall.

Olson, Mancur. 1965. *The Logic of Collective Action.* Cambridge, MA: Harvard University Press.

Ordeshook, Peter C. 1986 *Game Theory and Political Theory.* Cambridge: Cambridge University Press.

Ordeshook, Peter C. 1970. "Extensions to a Model of the Electoral Process and Implications for the Theory of Responsible Parties." *Midwest Journal of Political Science* 14: 43-70.

Page, Benjamin I. and Calvin C. Jones. 1979. "Reciprocal Effects of Policy Preferences, Party Loyalties and the Vote." *American Political Science Review* 73: 1071-89.

Palfrey, Thomas R. and Howard Rosenthal. 1985. "Voter Participation and Strategic Uncertainty." *American Political Science Review* 79: 62-78.

Rubin, Donald. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons.

Ruelle, David. 1991. *Chance and Chaos.* Princeton: Princeton University Press.

Ryu, Hang K. 1993. "Maximum Entropy Estimation of Density and Regression Functions." *Journal of Econometrics* 56: 397-440.

Sanchez, Maria Elena and Giovanna Morchio. 1992. "Probing 'Don't Know' Answers: Effects On Survey Estimates and Variable Relationships." *Public Opinion Quarterly* 56: 454-74.

Schulman, Mark A. and Gerald M. Pomper. 1975. "Variability in Election Behavior: Longitudinal Perspectives from Causal Modeling." *American Journal of Political Science* 19: 1-18.

Shannon, Claude. 1948. "A Mathematical Theory of Communication." *Bell System Technology Journal* 27: 379-423, 623-56.

Shepsle, Kenneth A. 1972. "The Strategy of Ambiguity: Uncertainty and Electoral Competition." *American Political Science Review* 66: 555-68.

Stigler, Stephen M. 1986. *The History of Statistics: The Measurement of Uncertainty before 1900.* Cambridge, MA: Harvard University Press.

Teixera, R. 1992. *The Disappearing American Voter.* Washington: Congressional Quarterly Books.

Tribus, Myron. 1961. "Information Theory as the Basis for Thermostatics and Thermodynamics." *Journal of Applied Mechanics* 28: 1-8.

Tribus, Myron. 1979. "Thirty Years of Information Theory." In *The Maximum Entropy Formalism*, eds. Raphael D. Levine & Myron Tribus. Cambridge: MIT Press.

Van Campenhout, Jan M. and Thomas M Cover. 1981. "Maximum Entropy and Conditional Probability" *IEEE Transactions on Information Theory* 27(4): 483-9.

Yatchew, Adonis and Zvi Griliches. 1984. "Specification Error in Probit Models." *Review of Economics and Statistics* 67: 134-9.

Wattenberg, Martin P. 1990. *The Decline of American Political Parties.* Cambridge, MA: Harvard University Press.

Wolfinger, Raymond E. and Steven J. Rosenstone. 1980. *Who Votes.* New Haven: Yale University Press.