Testing Theories with Qualitative and Quantitative Predictions

Coleman, Stephen

Metropolitan State Univ. (retired)

10 September 2005

Testing Theories with Qualitative and Quantitative Predictions

Stephen Coleman[1]

Presented at the General Conference, European Consortium for Political Research, Budapest, 10 September 2005

Researchers in the social sciences as well as other disciplines rely on statistical models to develop and test theories. The standard approach applies multivariate analysis—usually linear regression—and statistical hypothesis testing to observational data. The researcher may have several independent variables in mind as candidate predictors of the dependent variable; those reaching statistical significance compose the final model. In other situations, a theory is assumed to be correct and regression analysis is used to estimate parameters. Though experiments with random assignment of subjects are recognized as the "gold standard" for research, these are rarely possible in political science. Instead we rely on statistical controls to overcome problems inherent in using nonexperimental data.

This paper suggests that our confidence in using statistical methods to construct theories is misplaced and that theory testing is more productive when we combine definitive theory-generated predictions with statistical methods. I begin with a review of problems in the current approach to statistical analysis, then give several examples of how prediction can be improved.

Failures in Theory Testing with Observational Data

Research in political science is rarely subject to the degree of scrutiny that would expose the weakness in its statistical underpinnings. But examples from medical research show the seriousness of problems inherent in multivariate regression analysis typical of political science.

For some years physicians prescribed hormone replacement therapy for postmenopausal women in the belief that it would reduce their risk of heart attacks. Millions of women took their doctor's advice and, I would guess, billions of dollars

[1]Email to: dr.stephencoleman@gmail.com

were spent on this therapy. This protocol was based on observational research extending back many years, and there was sound theory to back it up. After menopause, a woman's risk of heart attack increases, just at the time her estrogen levels decrease. Further, treatment with estrogen is known to increase good cholesterol and decrease bad cholesterol, which ought to reduce the risk of a heart attack. This theory and the use of hormone replacements were backed up by positive findings in over 30 statistically controlled observational studies. But no randomized experiment on this therapy had been done before the therapy was widely adopted. It was recognized that women who sought replacement theory might be healthier or better educated than women generally are, lowering their risks of heart disease. But researchers tried their best to control these possibilities statistically.

This entire therapy was proved unfounded by Manson et al (2003) when results from several large experiments came in. The Women's Health Initiative, a randomized experiment involving over 16,000 postmenopausal women, found that this therapy actually increased the risks of coronary heart disease. The observational studies had mistakenly validated a wrong theory with the outcome that many women may have developed heart problems as a result of treatment, and a tremendous amount of money was wasted on ineffective treatments.

This example should be a great caution to political scientists, as we are working with less well developed theories and with less rigorous testing of our theories than the medical community. This is not an isolated example of a failure of observational studies in the medical area. Let's look at another case.

Over the past few decades we have gotten a lot of advice on what we should eat or not eat because of its supposed health benefits or risks. Almost all of this has been based on statistical analysis of nonexperimental data, and much of it has proved wrong when subjected to experimental research. For about 10 years, it was widely believed that Vitamin E might protect us against cancer or heart disease by mopping up destructive free radicals circulating in our bodies. In 1993 the New England Journal of Medicine reported two observational studies, statistically controlled, that showed a decline in heart disease among people who took extra amounts of Vitamin E. One of these studies (Rimm, et al. 1993) involved almost 40,000 men who were observed over a period of four years—a study that by its size and duration would put it among the best ever done if it had been done in political science. Again, however, recent experimental evidence has contradicted the efficacy of Vitamin E. A meta-analysis of experimental studies with a total of 80,000 subjects published in 2003 by Vivekananthan et al. found no benefit for Vitamin E in heart disease. So we have another example of failure in very well done nonexperimental studies with good statistical control and solid theory.

The failure in the early vitamin E research and similar nutritional studies is likely a result of model uncertainty, which can bias results (Chatfield 1995). Statistical testing is based on the assumption that the model in known. Model uncertainty arises when the model is not specified but based on the predictors that best fit the data. For example, one may include ten independent variables as candidate predictors and then use only the four that are statistically significant for the final model. Model uncertainty also comes in when one is coding variables into categories, as well as with nonlinearities and interaction among variables. The model produced may overly depend on extreme values in the sample, which biases the result toward lower p values and smaller standard errors. Because the final model depends on randomness in the sample, the model selection process is itself a random data-dependent process.

Much of contemporary statistical research is a ‘junkyard of unsubstantiated confidence’ based on false positive results (Longford 2005: 471). The rate of failure has been estimated by Ioannidis (2005) in a review of 49 highly cited medical research articles from 1990 to 2003—each cited over 1,000 times. Among the 45 original articles that claimed to find effective interventions, 7 (16%) were contradicted by later research. This still underestimates the false positive rate for nonrandomized studies, however; 5 of 6 were contradicted by later research. Among randomized studies 9 of 39 were contradicted.

One might conclude from these examples that we should abandon nonexperimental studies. I don’t believe this is necessary, though I am an advocate of political scientists doing more experiments and looking for natural experiments. The natural sciences have succeeded in part by developing theories that make definitive predictions—predictions that would almost certainly not be found if the underlying theory were false.

Would better statistical methods solve the problem? Statisticians are working on the model uncertainty problem, but a solution seems distant. Meanwhile, social scientists and econometricians have devoted a tremendous amount of work over the past several decades to the improvement of statistical methods. It seems fair to consider whether this effort has produced better social theories. So let's review some critiques that economists have made of research in their field. This is pertinent in that political scientists seem to draw more and more on econometrics to solve problems with multivariate analyses.

<u>Econometrics and Economic Theory</u>

For the past half century or more, from the time of the Cowles Commission, economists have worked diligently to create an economic science combining theory, measurement, and statistical analysis. The field of econometrics is impressive testimony of this. In the 1990s, however, a number of economists began to express their doubts about the success of this endeavor.

Economists have the same research standards as political scientists. To Lawrence Summers (1991: 129) "the best empirical work in macroeconomics formally tests substantive hypotheses rigorously derived from economic theory [and] sophisticated statistical techniques can play an important role in sorting out causation in systems with many independent variables." But his conclusion from reviewing progress in economic science is that this method almost always fails to find causal relationships that are not already obvious from the start. In his view, econometrics has had a negligible impact on the development of an economic science. Replication is mostly absent and rarely important, and methodological advances are confused with substantive progress. As to statistical testing he comments (p. 135) that "without some idea of the power of statistical tests against interesting alternative hypotheses . . . statistical tests are uninformative." He asserts that the best empirical evidence needs the least complicated statistical support and tells its story without much concern for the specific method of analysis.

Theory testing in economics went well up to the 1970s. Models derived from theory were mainly linear or had simple nonlinear forms. They were easily tested with regression analysis to identify relevant variables that should be included—albeit a very limited expectation for testing a theory. This situation changed in the 1970s, however, as empirical weaknesses in the old models became more apparent and the methods were not suited to newer dynamic models (Pesaran and Smith 1995; Heckman 2000.) Pesaran and Smith cite (p. 76) a yet more fundamental problem: "there is no agreed on method of judging whether the conditional predictions match the data and thus whether the evidence rejects the theory." If the data do not fit the theory, one cannot be sure if one should reject the theory, the methods used, or auxiliary assumptions made to test the theory. In their view, these conditions now make theory testing almost impossible.

The inability to test theories also calls into question the practicality of Popper's idea of theory falsification as the path to scientific progress. Dharmapala and McAleer (1996) contend that often the truth of an economic theory is assumed. The goal of empirical testing, as it often happens in everyday research, is the estimation of unknown parameters in a model, not a test of the theory itself. If there is more than one theory for a set of phenomena or more than one model can be derived from a theory, econometrics is not much help. If the criteria for falsification are too strong, one may reject all theories; if the criteria are relaxed, however, then one cannot reject the false theory with certainty. From their analysis one can interpret that the falsification ideal is subject to an arbitrariness that makes it impractical as a guide for scientific progress.

These commentaries on the state of economics resonate with the problems political scientists face. So how do we find our way out of the forest? The methods of physics are one source of advice. I suggest that we also examine how people construct theories in everyday life. Psychologists have recently come to a good understanding of the cognitive processes in our naïve construction of theories about the world around us. The give us an exemplar to consider when setting a standard for what makes for good testing of a theory.

Psychology of Theory Construction

To understand how the world works, humans develop and test theories as a normal part of life. Our ability to do this is innate and follows a process that psychological research is beginning to reveal (Ahn et al. 1995; Cheng 1997; Buehner, Cheng and Clifford 2003). Earlier, psychologists believed that people used information about covariation to identify the causes of events. Supposedly, people were like naïve social scientists; they observed what factors were correlated with an event and then constructed a theory about cause and effect. But new research refutes this idea. Instead, people have in mind causal models and seek information that might confirm or reject them (Ahn et al. 1995).

If a person were using a covariation model to explain an event, for example, they would be looking for factors that are either present or absent when the event occurs. By contrast, in assessing a causal model a person is trying to elaborate a mechanism that might have produced the event, seeking out information for or against a range of possible theories. Research confirms that people use this cognitive approach.

Further, people evaluate whether a mechanism causes an event by thinking about the difference in probability between getting the event when the prospective cause is present less the probability of getting the event when the factor is absent. New research also shows, however, that this model must be weighted inversely by a factor that attenuates the expectation of a cause when alternative causes are more often associated with the event in question (Buehner, Cheng and Clifford 2003).

If we apply this psychological model to how we test theories in social science, it is possible that our audience will find social theories more convincing and useful. To do this, one should try to explain a phenomenon by starting with one or more specific theories that would explain it; then look for a situation or test for a specifically predicted result that has a high probability of being associated with the prospective theory but otherwise is very unlikely to be observed.

By the psychological standard, one can see problems with postdictive regression analyses. Regression models often are not derived from a strong or specific theoretical base but are used to explore for possible explanatory variables. Alternative regression models often are available for the same phenomena, and the models produced by a regression analysis are often consistent with multiple theories of the same or different phenomena. In everyday life, people do not look to covariation to explain events, and we should be cautious about doing that in social science.

Examples of Predictive Theory Testing

The physical and biological sciences are good sources for ideas on what types of theoretically derived predictions lead to better theory testing. Here are several examples:

- Prediction of a constant or invariant (like the speed of light or a freezing point)
- Prediction of a specific number (typical of weather forecasting)
- Prediction of a symmetry, often derived from a mathematical model
- Prediction of a topological fixed point
- Prediction of a limit or constraint, or dynamic limit cycle
- Prediction of a specific or unusual dynamic behavior pattern
- Prediction of a specific spatial (geographic) pattern
- Prediction of an unusual statistical distribution
- Prediction that data will have a "signature" –- a unique mathematical shape (as used for detecting heart arrhythmias, nuclear tests, tsunamis, or submarines)

Sometimes a predicted result is better tested or more evident when the data are transformed and displayed in a different coordinate system. I'll give a few examples below.

Unusual statistical distributions are getting more attention for the qualitative prediction or rare or extreme events, such as nuclear accidents, financial crises and stock market crashes, natural disasters, ecological changes, and violence or wars (Sornette 2002; Coles 2001; Malinetskii and Kurdyumov 2001; Clauset and Young 2005). Extreme events follow a power law type of statistical distribution that makes them more likely to occur than one would expect from a normal distribution. To estimate the underlying distribution one samples the tail of the distribution rather than, for example, the central tendency or variation. Obviously these are important phenomena for policy makers to understand, whether they relate directly to political events or affect governments through the scale of disaster.

We can add a few examples to this list from game theory, such as predictions that people will tend to make a certain decision given the rules and payoffs of a game (Smith 2003); or from psychology, that people usually respond in specific ways to information, their surroundings, or what other people are doing (Plous 1993). Much work needs to be done, however, to understand how these individual decision processes play out in a large society.

Predictions about voting behavior using the entropy measure

To demonstrate my argument I would like to give several examples of both qualitative and quantitative theory testing using specific predictions. I am drawing these examples from my own research over many years and refer the reader to the original publications to get the full basis for the predictions. I only sketch some of the theory and analysis because my aim is to show a variety of approaches while limiting the substantive detail presented on each.

The examples I use are all from voting behavior and use the same quantitative measure, the Shannon entropy measure of information theory. The entropy measure is widely used in many disciplines, has well-known mathematical properties, and has a deep connection to human cognitive processes. It has been used by political scientists to describe political party systems much like Professor Taagepera's measure for effective number of parties. [See Kirchgaessner and zu Himmern (1997) for an example of the use of entropy to analyze German elections.] Other mathematical measures might give similar results, but I have not investigated that.

Given a set S of n alternative events or choices, with probability p_i for i = 1, 2, . . . n, the entropy of the set is

$$H(S) = - \sum p_i \log(p_i)$$

Entropy is always positive and attains its maximum value log n when all events are equally probable. Usually the logarithm is to base 2 and the units are bits, a convention followed here.

Among other possibilities, entropy can be used as a measure of dispersion of votes across the party system in parliamentary or American presidential elections. The entropy of the distribution of votes across political parties can be calculated from the portion of votes each party wins or, equivalently, from the probabilities that randomly chosen voters vote for each party. In a two-party system, entropy will be greatest when the party vote shares are closest to equality and the election is most competitive. As one party becomes dominant, entropy decreases. In a multiparty system, entropy increases as the vote share shifts from the largest parties to minor parties, as when the party system becomes increasingly fragmented or new minor parties draw voters away from older parties. This may or may not represent a more competitive election. Entropy generally increases with the number of parties.

Let H(P) be the party entropy calculated retrospectively from the fraction of the vote going to each party. For example, if the vote divides among three parties in the proportions 0.60, 0.30, 0.10 then

$$H(P) = - \sum p_i \log p_i = - .6 \log .6 - .3 \log .3 - .1 \log .1 = 1.29 \text{ bits}$$

In party entropy calculations here I used the number of parties as typically presented in election data by the sources that report the data. If the data included an "other" parties or "minor" parties category, I added that as another category in the entropy calculation. The number of parties is a fuzzy measure and I didn't want to introduce my own bias as to what to count. If one is only speaking of parties in a legislature, then the number of parties may be well determined. But when people are voting, they may differ in their awareness of the competing parties; the number of parties depends on how most people categorize them perceptually. If people are not aware of a party, then it will have no bearing on their voting decisions. So, I would say that if people are not aware of some minor parties or lump them together conceptually, that is how they should be dealt with analytically when analyzing voting data. This is a problem that can be investigated empirically, of course.

Qualitative predictions

Cyclic Behavior. Cyclic behavior is common in systems of all types. Business cycles, for example, are frequently observed and are a topic of long-standing interest in economic theory. A similar question comes up in elections. Do fluctuations in the vote shares for parties follow a regular pattern? Is there a tendency for a two-party system to cycle around a midpoint? Is there a stable equilibrium in two-party or multi-party systems?

The relatively stable party systems in most established democracies suggest that voters are interested in stability. Suppose, as a hypothesis, that voters try to maintain a fairly stable equilibrium in their political party system, such that no one party completely dominates the others in all elections and that minor parties are not eliminated. If one party is getting too far ahead of the others in this scenario, some voters may react by voting for an opposition party. It is known that any system controlled by negative feedback will oscillate if there is a delay in the feedback loop. That is, if to maintain stability some people adjust their voting behavior in response to results of the last election or, perhaps, expected results of an upcoming election, and they are not aware that some other voters are doing the same thing, there likely will be an over-correction in voting. Continued voting decisions like this will produce cycles in the party system. This is one instance where theory may predict cyclic change.

So how should one look for cyclic behavior in a party system? Suppose that we use the entropy measure or a similar mathematical measure that captures the overall distribution of party vote shares to assess cycles in voting. Typically one might think first of a time series analysis, but this can be difficult with small samples and it invites many statistical problems. Instead I suggest changing the coordinates to makes an easier, qualitative test of the theory.

To study change one can use a special graphical technique also known as a Poincare phase space map. I plot political party entropy $H(P)$ and its rate of change in each election successively on a graph and connect the points. The horizontal or x-axis represents the value of $H(P)$ at an election; the vertical graph is the rate of change of $H(P)$ at that election. Change in $H(P)$ at a given election is the amount of change from the previous election, that is, the difference between the current value and the previous. Because we only have measurements of $H(P)$ at elections, the graph is a series of points, not a smooth or continuous line, but we can connect the points with straight lines to show the path of change. If the data have a trend as well as a cycle, the trend should be removed before analysis.

This type of graph is well suited to showing change that cycles, including cases where the system comes to an equilibrium and no further change occurs for a time. If change is cyclical, like the motion of a pendulum, the graph of change is approximately an ellipse or circle with the system moving clockwise around the ellipse. See, for example, Figure 1. If the system comes to rest, like a pendulum gradually slowing to a stop, the graph will be a clockwise spiral in toward the equilibrium or resting point. As a system cycles around, one can easily compare one cycle with the next, as the path of the system overlays itself repeatedly on the graph.

One must be cautious, however, in interpreting causality from an elliptical pattern. A sequence of random changes can also look cyclic or quasi-cyclic. For example, when counting ballots in a close election between two candidates, one candidate may be ahead for a time, but then the other candidate may take the lead, and then the first candidate again, and so on, with the lead switching from on to the other. (This pattern is known as the arc sine law.) One can try to distinguish between a random cycle and a causal cycle with statistical tests, though these may lack power if the sample is small. For example, if there is a correlation between consecutive points, this would cause one to reject randomness. One can also use a nonparametric runs test. To my knowledge, political science does not have a theory of random change in voting for political parties that would produce cyclic behavior in the party system. Economists, by contrast, have discussed this quite a lot in analyzing stock market prices.

Now let’s apply this method to parliamentary elections in postwar France from 1946 to 1988. Data are from Mackie and Rose (1991). Figure 2 shows first the conventional time series view, plotting party entropy against the year of the election. Figure 3 shows the data in the Poincare phase space coordinates. To my eye, Figure 3 brings out the cyclic behavior much more clearly than Figure 2. Statistical tests, however, do not reject the possibility that this change is random, but the power of the test is small with a sample size of only 13 elections. (Or one might say that the results are inconclusive as to randomness.) I have found similar cyclic patterns in voting in a number of countries, though the rate and amplitude of the cycles varies across countries (Coleman 1993).

Figure 1. Graph of clockwise, cyclic change in entropy.

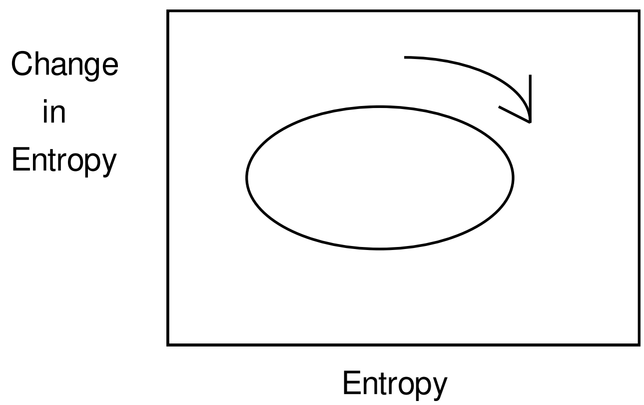


Figure 2. Party entropy change in postwar France (1946-1988)—time series.

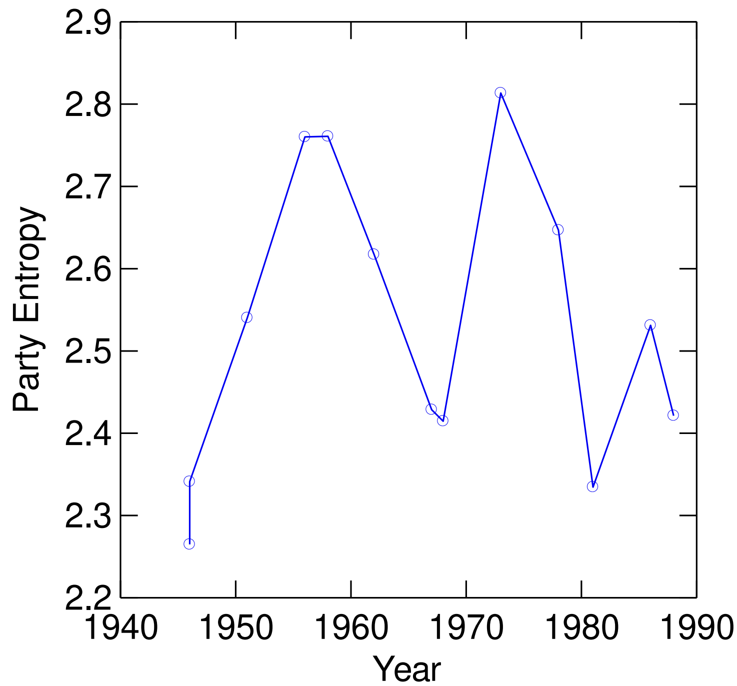
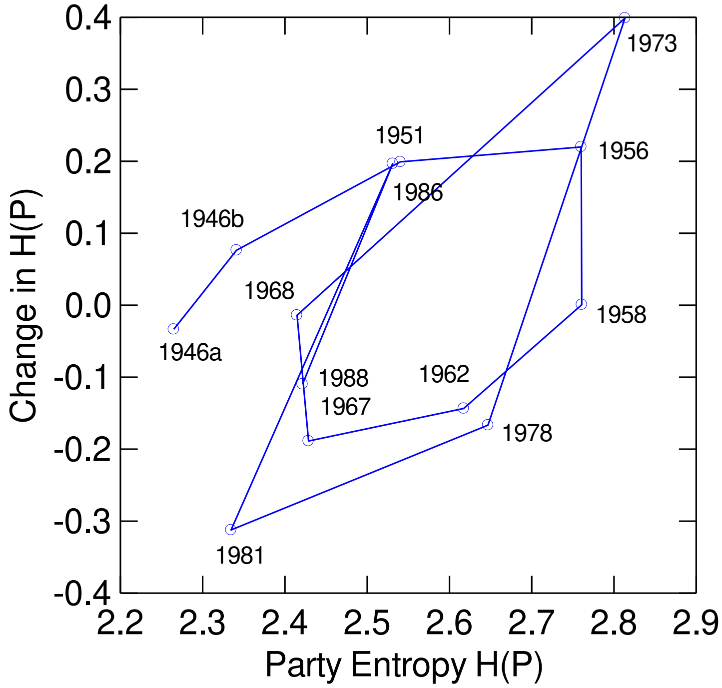


Figure 3. Party entropy change in postwar France—Poincare phase space.



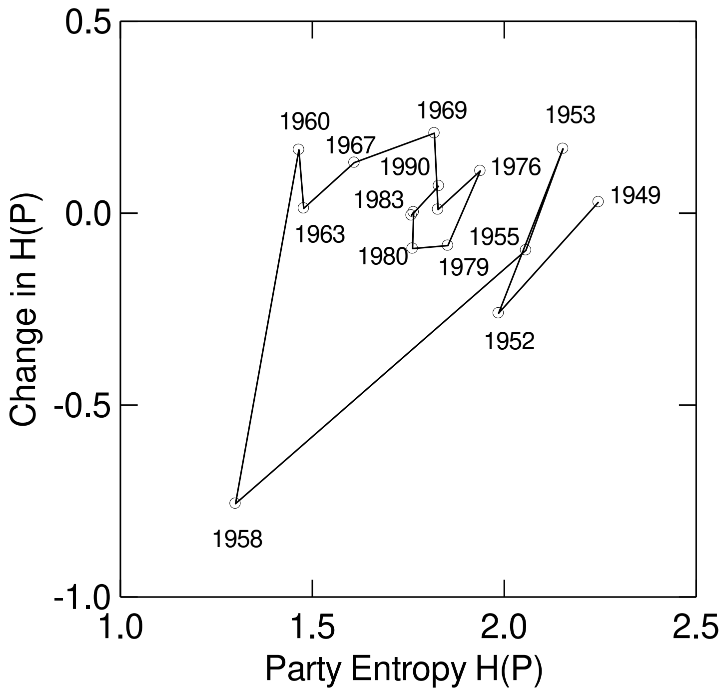
A good example of a political system cycling toward an approximate equilibrium is Japan in the 17 postwar elections from 1947 to 1990, as seen in Figure 4. The data for each election are national averages for the 46 prefectures (excluding Okinawa). (Japanese election data are published in annual volumes of the *Japan Statistical Yearbook* in English language.) The graph shows that the party system starts out with some noncyclic, jerky changes in the first few postwar elections, then makes a wide cycle of change closing in to smaller cycles after 1967; party entropy changes very little from 1972 to 1990, when the system is close to equilibrium. Statistical testing rejects randomness for this case. One cannot determine, however, if the equilibrium is stable.

We can also examine dynamics and equilibrium in a different coordinate system, which some also call phase space analysis. We put $H(P)_n$ on the horizontal axis and $H(P)_{n+1}$ on the vertical axis for election n and party entropy H(P).

An equilibrium satisfies the condition that $H(P)_{n+1} = H(P)_n$ which we can represent with the 45 degree line when we plot $H(P)_{n+1}$ against $H(P)_n$, as in Figure 5 for the same sequence of Japanese elections seen in Figure 4.

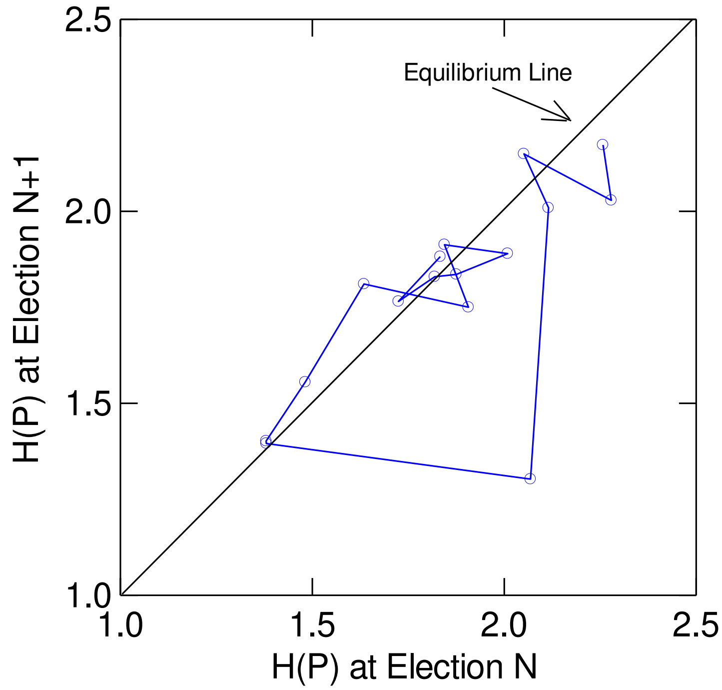
These examples are designed to show how changes in the coordinate system can improve our understanding of a dynamic process. These alternate representation of reality are especially useful, I believe, when dealing with short time series—say, fewer than 20 points—or when comparing one country with another—typical applications in political science. In these situations, I suggest, the gross qualitative characteristics of the dynamic patterns are the important features. Little more would be added by reformulating the short time series quantitatively, as in a Fourier analysis or ARIMA model. The key features of the graphs might be further enhanced, however, with smoothing techniques.

Figure 4. Party system change in postwar Japan approaching an equilibrium.



Note: The 1972 election is not labeled and the 1986 election, also unlabeled, is at the same point as 1983.

Figure 5. Another look at the equilibrium in postwar Japanese elections.



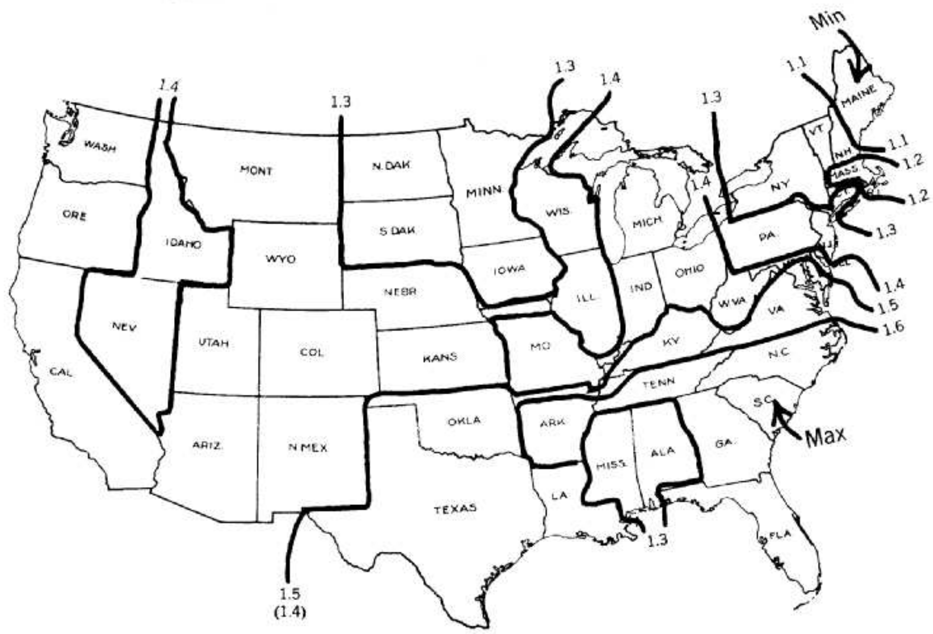
Harmonic distribution. Suppose that within a country people in each local geographical area are influenced in their attitudes or behavior by the people in immediately surrounding areas. More specifically, suppose that in any small area people's behavior is approximately an average of the behaviors in surrounding small areas. If the neighboring community to the west is high on some measure, and the community to the east low, then the community in the middle has a mid-level ranking. Now what would this look like if we extrapolate this to a large geographic area, an entire country?

The result is a spatial distribution with a unique mathematical form, the harmonic function. We are all very familiar with this in a different guise—the daily weather map that shows bands of equal temperature across a country or continent. Typically the coldest areas are to the north and temperature bands show gradual warming as one moves south. Other properties of harmonic functions are that the highest and lowest values will be on the border of the system, and the average around any circle will equal the value at the center.

As a fairly good example of a harmonic distribution, Figure 6 shows political party entropy contours across states in the US presidential election of 1968, which was a three-party election (Coleman 1975). Party entropy changes smoothly as one moves along the eastern seaboard from the highest entropy state, South Carolina, to the lowest, Maine. All the entropy contours in units of a tenth bit are crossed. The distribution is not perfectly harmonic but is supportive of the underlying theoretical model of spatial influence in voting. It would be difficult to get this result with any quantitative statistical method. The graphical result can be supported, however, by a conventional statistical analysis of correlations between neighboring states.

Although data analysts are well aware of serial correlations in time series data and how to deal with them, analysts seldom apply the same rigor to correlations in cross-sectional analyses of within county data. This is admittedly a difficult problem and spatial analyses, themselves, are much more difficult than regression models. But this example shows how simple qualitative mapping techniques can give us some insight to when spatial correlations should be considered in our analyses.

Figure 6. A near harmonic entropy distribution in the party vote for presidential candidates in the U.S. in 1968.



Quantitative predictions

The next examples are from my research on the effects of social conformity on a variety of collective behavior patterns, including crime and voting, which look very much alike from this point of view. My concern is the sociological or, better, the social-psychological influences of conformity or group pressure on voting behavior and how this may look on a large scale. Because there may be many types of voting behavior—people make their voting choices for different reasons—one must figure out how to detect conformist behavior and, if possible, distinguish it from other types of decisions. Does conformist behavior have a unique signature that would allow us to say, "Yes, that's conformist behavior and, No, it is almost certainly not rational decision making or any other type that we know about"? A predictive test like this fits the psychological model of how people develop causal theories.

In most countries a large share of the electorate vote because of the social norm that good citizens should vote (Blais 2000, Knack 1992; Knack and Kropf 1998). This is an example of the effect of social conformity, which can be a response to a social norm or group pressure, or occur when people adapt their behavior to what most people are doing in a given situation. It is also known that, in the aggregate, when people respond to one type of social norm they are also more likely to respond to other norms. The more similar the norms, the more spillover of conformity from one norm to the next. Salience and priming also increase people's conformity with a norm. That is, if we remind people to vote, those most responsive to conformist pressures will be induced to conform. At the individual level, however, people vary in their response to conformist pressures and situational effects are significant.

I propose using voter turnout in a national election as a standard for measuring the conformity level in a country, at least for those people most responsive to conformity. It gives a snapshot of how successful a nation is at getting people to follow a widely accepted social norm. But turnout must be adjusted quantitatively to how people perceive it in terms of social conformity (Coleman 2004). Psychological research reveals that when people observe events with a probabilistic distribution their response is better described by the entropy measure. That is, there is a nonlinearity in how people react to perceived probabilities; a change in the probability of an unlikely event has more impact than a change of equal increment of probability change in a very common event.

So one can calculate a turnout entropy H(T) from turnout t as

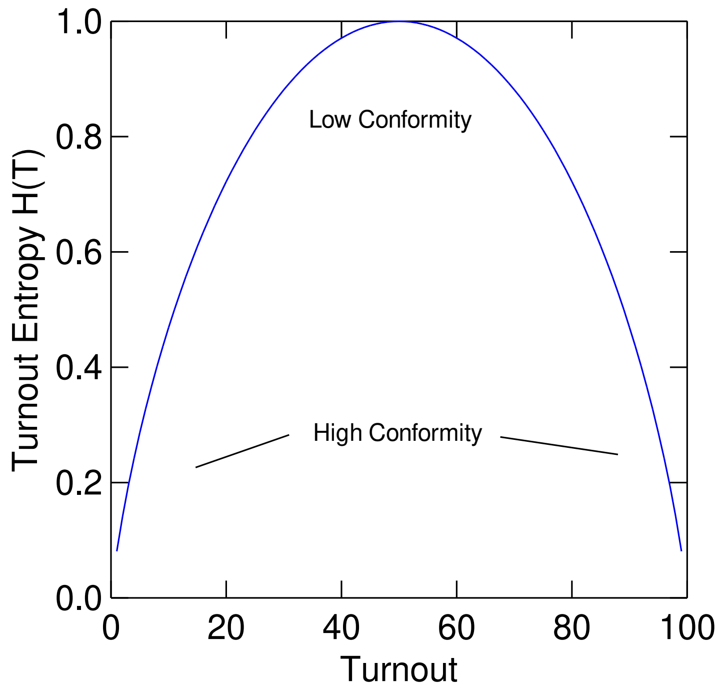
$$H(T) = - t \log t - (1-t) \log (1-t)$$

The theoretical relationship between voter turnout and perceived conformity, measured by entropy, is shown in Figure 7. Note that the least conformist situation is when the turnout is closest to 50%. This is where people would be most uncertain about whether the norm was to vote or not to vote. As turnout increases or decreases, those who are responsive to conformity will see a stronger norm toward voting or not voting. Note that this model assumes that for many people the act of voting involves only two choices, to vote or abstain.

This curve is a signature of conformity. It has distinguishing characteristics: nonlinearity approximately in the shape of a parabola, symmetry about the midpoint, and the maximum at 50% turnout. These are qualitative and quantitative features that we can use as predictions for testing the theory that social conformity affects voting or other social behaviors on a large scale.

We can predict from the properties of the entropy function (or Figure 7) that places with turnout closer to 50% also should exhibit lower social conformity on other types of social behavior that is typically moderated or controlled by social conformity (Coleman 2002, 2004). This is a quantitative prediction that one can test for, as well as the qualitative symmetrical relationship of conformity with respect to voter turnout above and below 50%.

Figure 7. The relationship between turnout, turnout entropy, and perceived conformity.



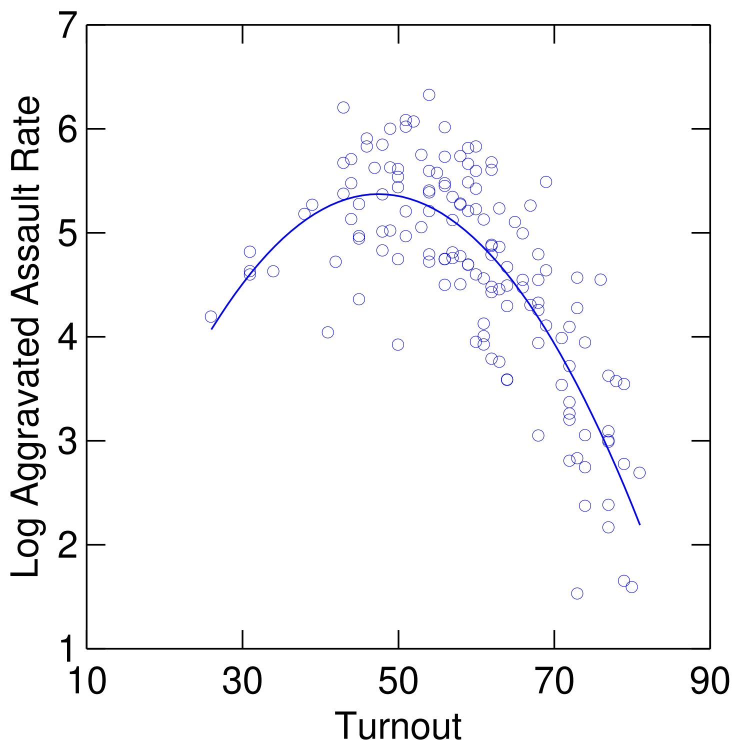
Crime Rates and Conformity

Let's examine first the proposition that conformity affects crime rate. It is well known that crime is partially controlled by social pressures to obey laws and social norms (Coleman 2002). In Figure 8, I show how the rate of aggravated (serious) assaults varies in relation to turnout in American states. I used crime data for three years (1960, 1970, 1980) matched with three presidential elections (1960, 1972, 1980). I picked these years because there were several states with voter turnout below 50%, which makes for a better test of the signature shape. I transformed the crime rate with the log function to facilitate additional statistical testing (as is often done with crime analyses). To test for a maximum at the predicted 50% turnout and the symmetry about that line, I fit a parabolic equation to the data as a function of turnout and estimated the location of the parabolic maximum from the coefficients in the fitted parabolic curve (also shown in Figure 8). That is, I am using statistical regression analysis to test a specific theoretical prediction, not in a hunt for possible causes or explanatory variables. To establish a confidence interval for the estimated location of the parabolic maximum, I used a bootstrap method, resampling the data with replacement 1,000 times and re-estimating the model each time. For added confirmation of the findings I did multiple regression analyses with independent variables known to be related to crime.

The net result is an estimated parabolic maximum in state aggravated assault rates at 46% turnout with a 95% bootstrap confidence interval (CI) of [36%, 51%]. The simple bivariate graphical analysis seen in Figure 8 holds up through the additional statistical control, and similar results are observed for other types of violent crimes and burglary (Coleman 2002). Although the maximum of the parabola is not exactly at 50% turnout, that point is within the confidence interval. I would also suggest that even if the parabolic maximum were not so close to 50%, these findings would still be strong confirmation of the theory based on the qualitative signature. There is no other good explanation that I can offer to explain the observed parabolic relationship between voter turnout and crime rate. Note also that the lack of causality makes this an interesting finding, as one does not expect that voting causes crime. Instead one must look toward a mutual explanation of both phenomena, such as conformity.

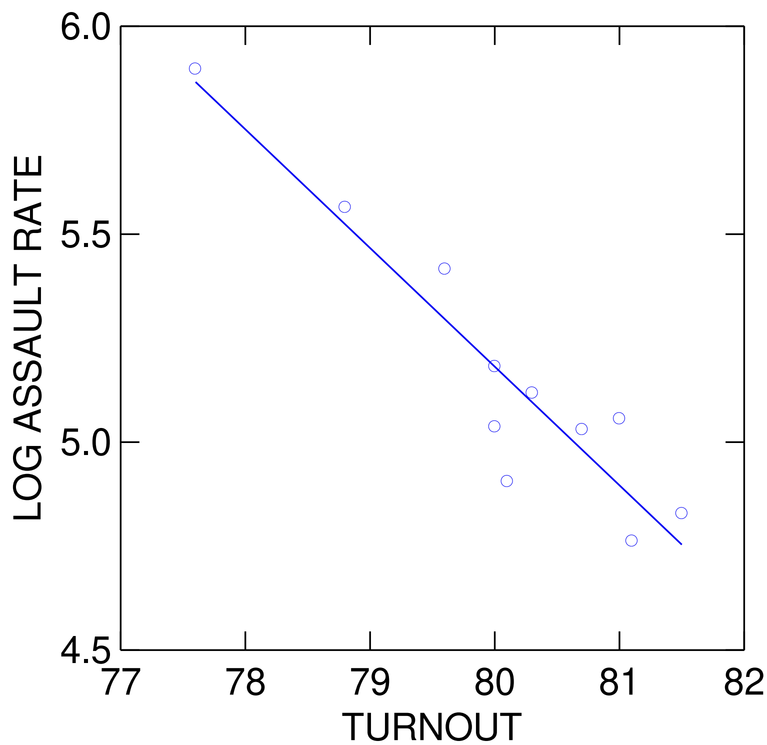
To confirm the generality of these results, I have looked at a few other countries. Figure 9 shows the same analysis for German Laender in the western part of the country. Because voter turnout in Germany, as in most European countries, has little within-country variation and is almost always over 50%, one sees only part of the parabolic curve—a part of the curve, in fact, that is more nearly linear than curvilinear. Serious assault rates in Germany fit the same pattern as in the US, though one does not see the parabolic effect in Germany. This is sharp reminder that one should be aware that the range of a variable may limit our testing of a theory and, if we are not careful, bias our results. (There are too few eastern Laender for analysis but they appear to have a similar pattern though at different levels of crime rates.)

Figure 8. USA; aggravated (serious) assault and turnout; 1960, 1970 and 1980 combined.



Notes: For the bivariate plot, N = 144, R squared = 0.61, parabolic maximum is estimated at 47% turnout.

Figure 9. Germany 2002; serious assault rates and turnout in the western Laender.



Notes: Crime data from Bundeskriminalamt, www.bak.de; N = 11; R squared = 0.87; p < .0001.

Spillover Effect. In earlier research (Coleman 2004) I showed that conformity in the decision to vote can spill over to affect voting choices for political parties. Psychological research informs us that people tend to be consistently conformist, and that if they are prompted to conform in one instance they are likely to conform in a different but related instance (Cialdini et al 1990; Knack 1992; Knack and Kropf 1998). So the degree of conformity one sees in collective voter turnout should have a correlation with voting for political parties. Not all voters may be affected by this, but we can test for the strength or prevalence of conformity in voting choices.

If a society is becoming less conformist, voters may shift to minor or new parties on the fringe while turnout shifts toward 50%. If a society is becoming more conformist, one might expect to see people shifting from minor parties to major parties, while turnout increases toward 100% or decreases toward 0. In extreme conformity one might see the suppression of minor parties and their adherents. Because party entropy is related to both the distribution of votes across parties and the number of parties, the prediction is that the number of parties likely increases as turnout shifts toward 50%.

One can express this idea of conformity spilling over from turnout to party choices more strongly and quantitatively. A first hypothesis or model is that entropy of party voting H(P) is positively correlated with entropy of turnout H(T); such as,

$$H(P) = a\ H(T) + b$$

Because the two entropy measures are on different ratio scales, each ranging from 0 to log k for k choices, we can rewrite this as a stronger hypothesis that

$$H(P) = \log k \;\; H(T)$$

In this model H(P) is predicted to be a multiple of H(T).

One can test these models directly—and I give some examples below—but a change in coordinates brings us to a better test of the predicted qualitative signature and symmetry. More often one would test a model with H(P) as a function of turnout instead of H(T). As a function of turnout, H(P) is predicted to look like the nearly parabolic curve in Figure 7 but multiplied by a factor of log k. I don't know of any other voting theory that would make a similar prediction.l

We can easily test this prediction by a conventional linear regression analysis, regressing party entropy H(P) on a parabolic (2nd degree polynomial) function of turnout t

$$H(P) = a_0 + a_1 t + a_2 t^2$$

If the coefficients reach statistical significance, one can determine if the parabola has a maximum or minimum and then estimate its location with respect to turnout. (The extremum is at turnout = $-a_1/(2a_2)$; a bootstrap estimate is used to find a confidence interval for the location, as with the crime analysis above.) If the fit is closer to a straight line than a parabola, the analysis will reveal that as well. The parabolic regression also permits testing the predicted hypothesis against alternatives where the maximum of the curve is shifted away from 50% turnout, something not possible when regressing H(P) against H(T).

Testing this hypothesis is essentially the same as the analysis of crime rates. One can look for a symmetrical distribution of party entropy H(P) centered near a turnout of 50%. This should be true in either longitudinal (time series) data and cross-sectional data. To make a good test of this, I needed to find elections with a wide range of turnout, extending well below 50% to higher levels. Fortuitously, turnout did vary widely across American states in U.S. presidential elections in the early decades of the 20th century including states in the South with very low turnout.

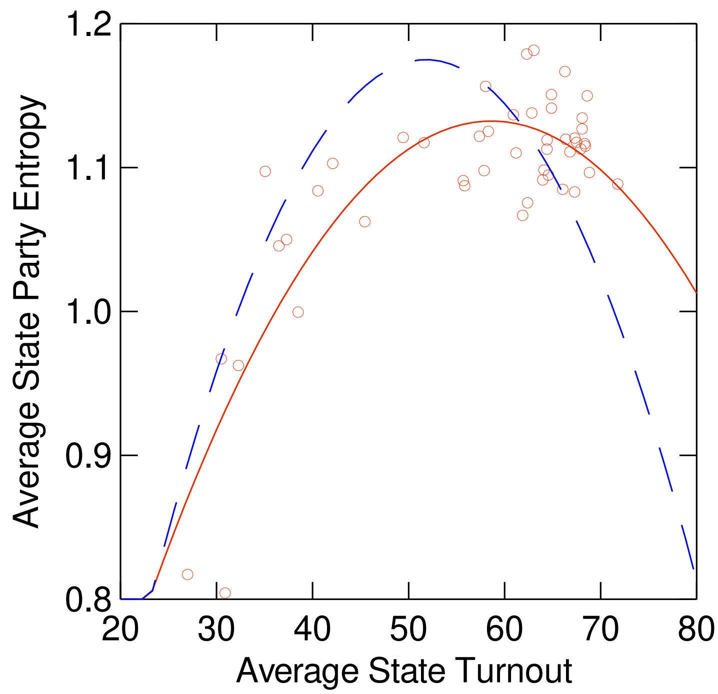
If one aggregates all American states and elections from 1920 to 1988, one finds a parabolic maximum at turnout = 56% with a 95% CI [54.6%,58.1%] (See Coleman 2004 for more detail; women's suffrage first covered all states in 1920.) There are too many data points to show this graphically, so as another example I averaged the turnout and party entropy over each state from 1920 to 1996 and fitted a parabolic curve, which has a maximum at 58.6% turnout (Figure 10); averaging, of course, increases the explained variation. To look at a single election, see Figure 11, the result for testing the prediction for the U.S. 1932 presidential election. In this case the maximum is shifted above 50% to about 65%. Does this departure from 50%

cause us to reject the hypothesis? I don't think so; but it might cause us to look for additional factors that may have caused the results to vary from the prediction in specific elections. For example, a mix of conformity with other types of voting behavior might change the outcome. Figure 12 shows the results for 1968 when George Wallace ran as a strong third-party candidate, especially in the South. Turnout increased in the South nearer to 50%, and the highest values of party entropy also are near 50 percent turnout as seen in Figure 12. Compared with 1932, one sees here more of the right half of the parabolic-like distribution.

See Coleman (2004) for more results and analysis for elections in the U.S., Western and Eastern Europe, and Russia. (For comprehensive results on Russia and Eastern Europe see also the ancillary material for *Political Analysis* Volume 12(1) (2004) on the Website of the Political Methodology Society, a section of the American Political Science Association, http://polmeth.wustl.edu/polanalysis/ancillary12.html.)

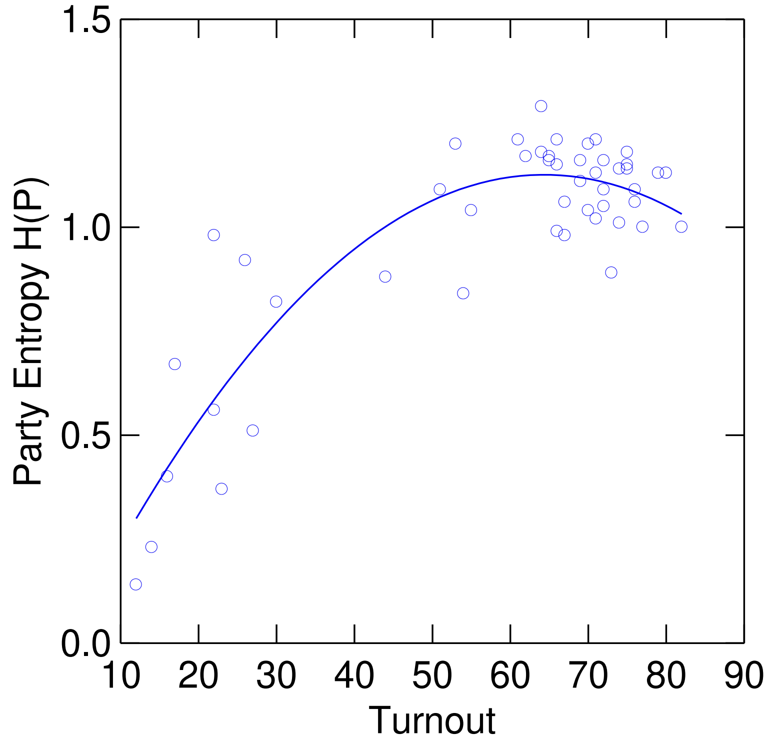
Figures 13-19 illustrate the results for several European parliamentary elections. I picked examples that show strong evidence of the predicted conformity pattern in voting behavior, including Russia, Germany, Austria, and Switzerland, as reported in Coleman (2004). In these cases one can easily see the qualitative aspects of the predicted model. Minimal statistical results are reported in notes to the figures, as the findings are already evident in the graphical presentations. Moderately strong conformity effects (not shown) are also found in Norway, Sweden, Spain, Romania, Ukraine, and Poland. Countries where the effect is weak include France and Britain, though Figure 19 is an exception, showing a moderate conformity effect on the British vote for the European Parliament. The analysis excluded countries with compulsory voting.

Figure 10. U.S. presidential elections, 1920-96; mean party entropy *H(P)* of each state in relation to its mean turnout in elections from 1920 to 1996; shown with parabolic curve fitting (solid line) and entropy curve estimated at 1.19 *H(T)* (broken line).



Notes: N = 48; parabolic fit R square = 0.72; maximum of parabola estimated at 58.6% turnout; women's suffrage began in 1920.

Figure 11. U.S., 1932, parabolic relationship between turnout and political party entropy *H(P)* for the 48 states.



Notes: Estimated parabolic maximum is at 65% turnout with 95% bootstrap CI = [59,75]; R square = 0.79; p < .0001.

Figure 12. Party entropy H(P) and turnout in US 1968 presidential with a strong third party; theoretical maximum party entropy for three parties is 1.58 at 50 percent turnout; data shown with linear fit.

Note: N = 48; R square = .45; p< .0001.

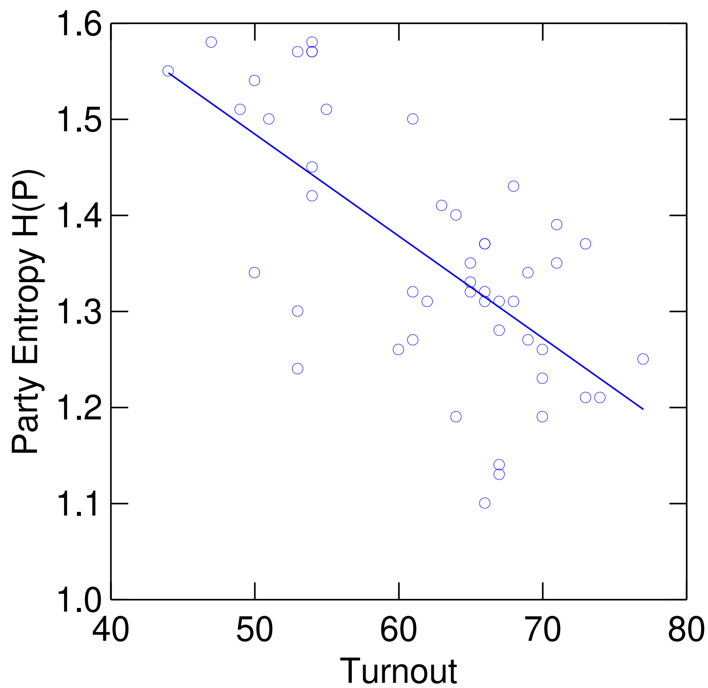
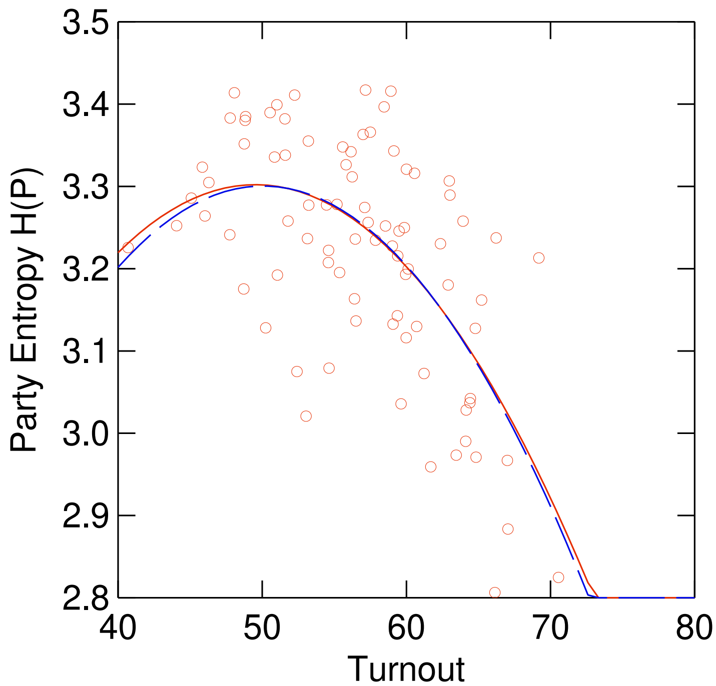
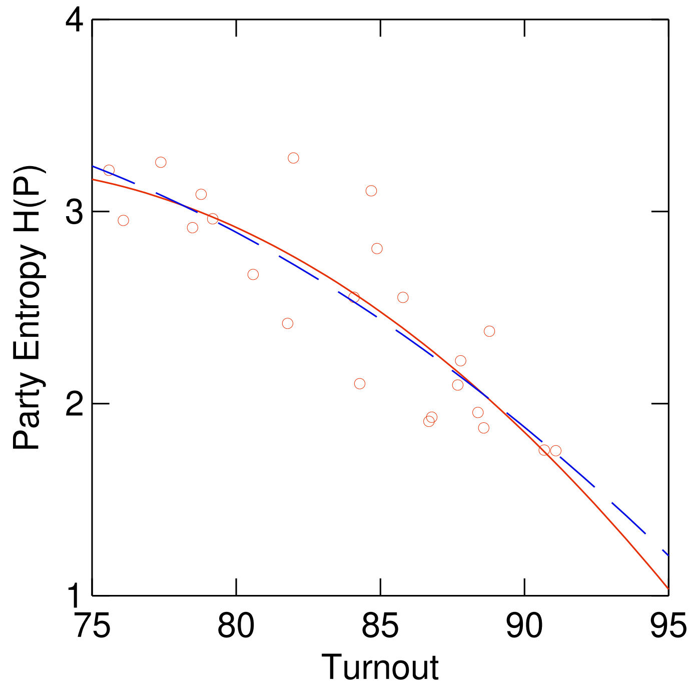


Figure 13. Turnout and party entropy in the Russian parliamentary election of 1993; district level analysis with parabolic curve fitting (solid line) and entropy curve (broken line) estimated by regressing on turnout entropy as in Figure 7.



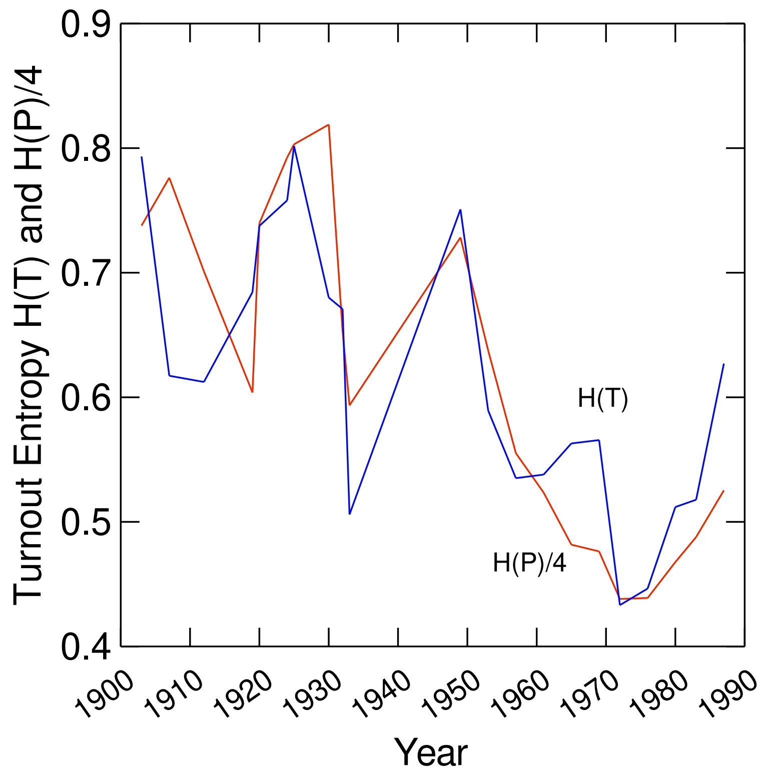
Notes: N = 84, one outlier removed; R square = 0.40; parabolic maximum estimated at 49.5% turnout with bootstrap 95% CI = [36.6, 51.9]; H(P) = 3.3 H(T).

Figure 14. Germany, 1903-87, turnout and political party entropy *H(P)*; data fitted with parabolic curve (solid line) and entropy curve (broken line) estimated by regressing on turnout entropy curve give almost identical results.



Notes: N = 23; R squared for linear fit is 0.70. Correlation of number of parties with turnout is -0.68 (p = .0003).

Figure 15. Germany, 1903-87, time series representation of turnout entropy *H(T)* and political party entropy *H(P)* show same proportional change in each measure.



Notes: N = 23; party entropy H(P) divided by 4 (= log 16) to align with H(T) based on regression estimate of model H(P) = c H(T).

Figure 16. Germany, 1903-87, linear regression of H(P) against H(T) also shows proportional change in each measure.

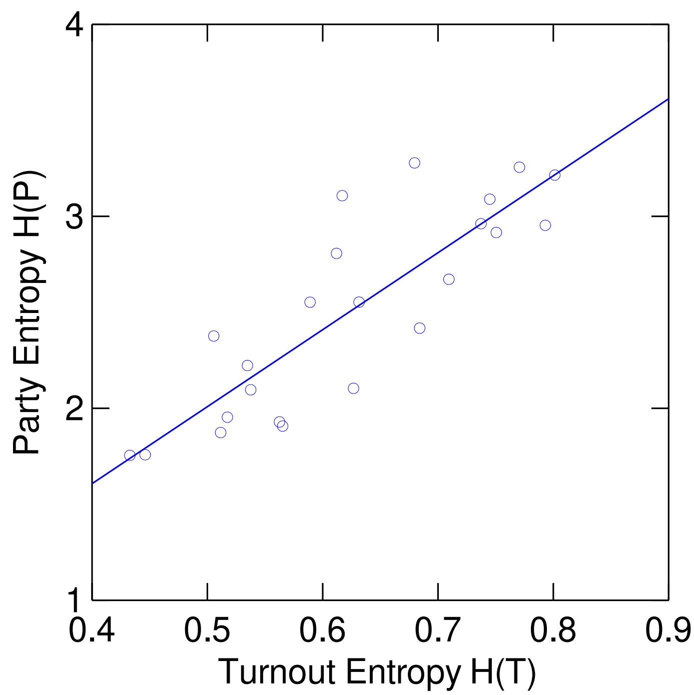


Figure 17. Austria, Germany and Switzerland: combined parliamentary elections from 1946 to 1987; relationship between party entropy and turnout.

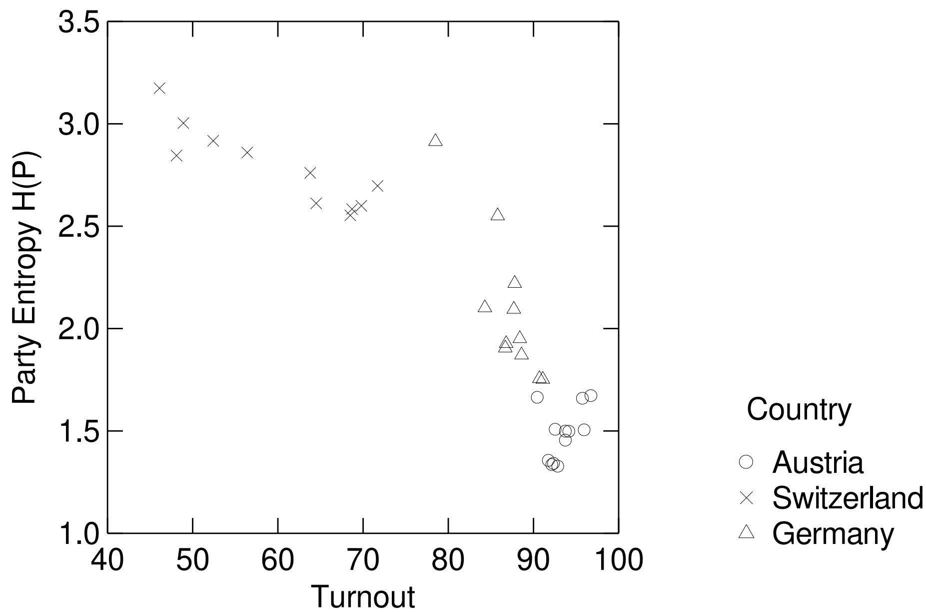
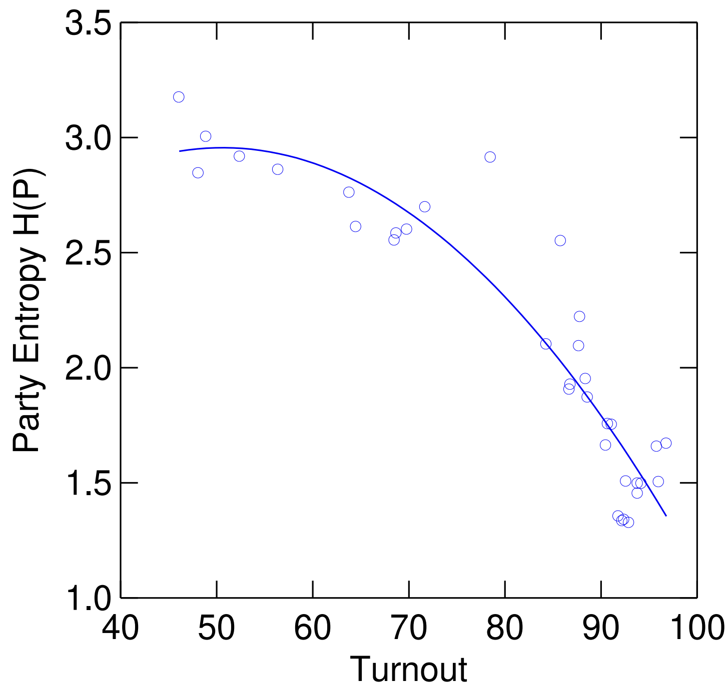
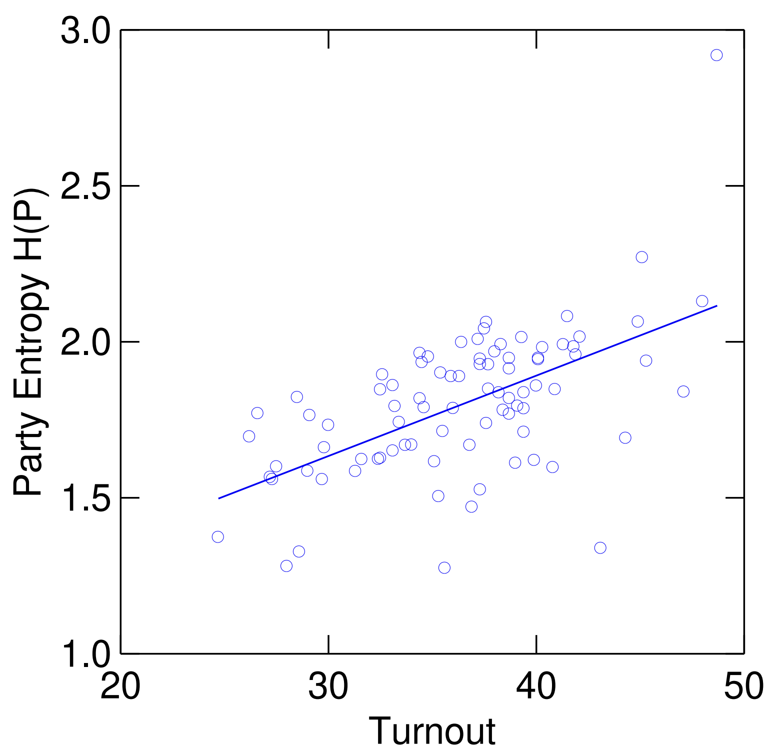


Figure 18. Austria, Germany, and Switzerland: combined elections from 1946 to 1987, relationship between party entropy and turnout with parabolic curve fitting.



Notes: N = 34. For parabolic fit R square = 0.87 and the parabolic maximum is estimated to be at 51% turnout.

Figure 19. U.K. election to the European Parliament, 1994, turnout and political party entropy *H(P)* with linear fit—a rare European election with turnout below 50 percent in all constituencies shows only one side of the expected parabolic relationship.



Notes: N = 83; R squared = 0.35; p < .0001.

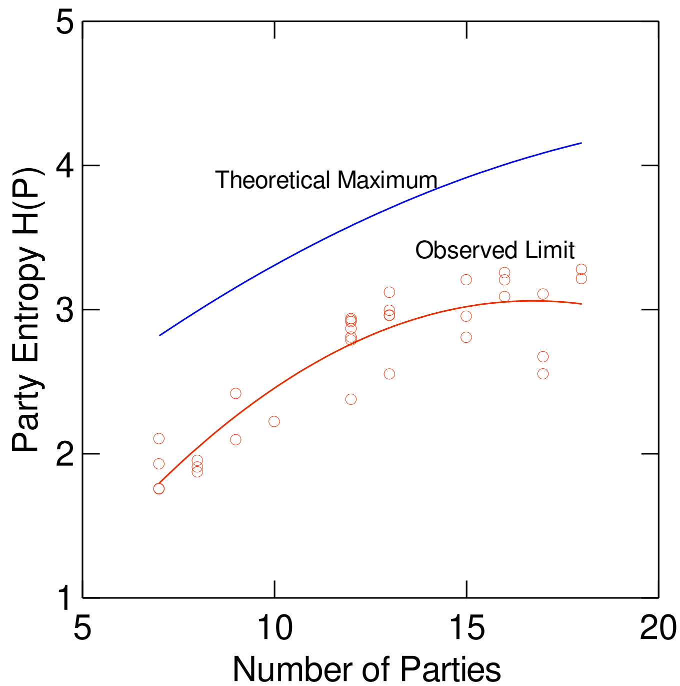
Constraints. As a final example I show how one might examine constraints and extremes in voting behavior. Psychological research shows that people have a definite limit on their capacity to process probabilistic information about categories along a single dimension (Miller 1956). This also seems to apply to group decision making, as when people are using and sharing the same probabilistic information (Miller and Bieri 1963). On the entropy scale this is about 3 bits, which corresponds to a person trying to accurately discriminate among 8 alternative choices or events if they are equally likely to be observed. A person can keep track of more choices if some are less probable than others. Near the psychological limit, people make more errors in categorization.

Researchers have also explored how consumers make choices and, as with the entropy studies, report inherent cognitive limits. Iyengar and Lepper (1999) showed this experimentally by offering consumers either six choices of jams or two dozen choices. When offered the higher number, consumers were much less likely to buy any; they preferred having the smaller number of choices. This experiment echoes research by Hauser and Wernerfelt (1990) showing that when consumers are faced with more options and more information about them, they tend to consider fewer choices and use less of the available information.

Assume that people are using and sharing probability information about political parties—the likelihood of other voters voting for each party or, equivalently, the percentage of votes likely to go to each party—when they decide how they will vote. Then one might expect to see a constraint on their collective behavior when measuring the political party entropy. One might also expect that voters will not tolerate too great a number of parties. To test these ideas, I examined the number of parties and party entropy H(P) for parliamentary elections in Germany and France from the late 1800s. Figure 20 shows the theoretical maximum party entropy in relation to the number of parties and the observed entropy in Germany. Note that the observed values increase with the number of parties competing but then level out and apparently decrease at the highest number of parties. The same holds for France (Figure 21). This pattern is similar to the behavior of consumers who have too many choices.

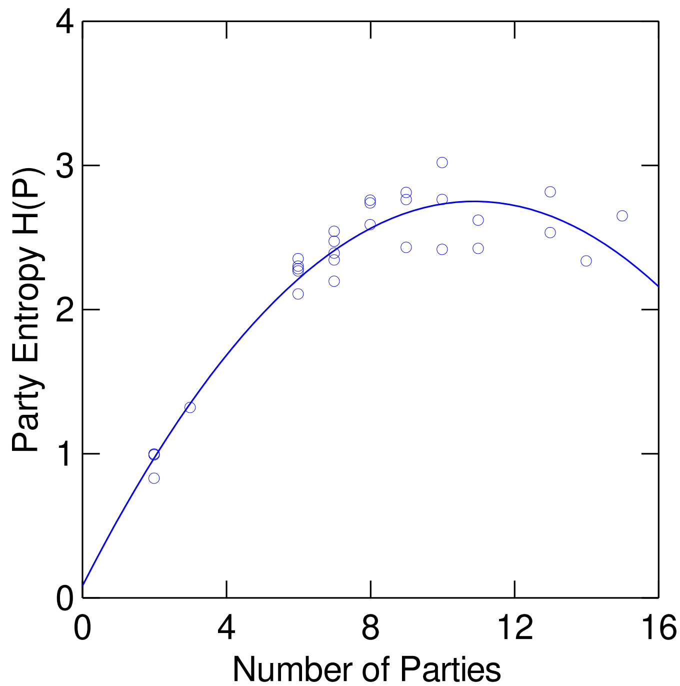
We can supplement the graphical analysis with statistical analysis. For this period in Germany there were 33 parliamentary elections. Party entropy ranged from 1.7 to 3.3 bits. Assuming that this is a representative sample of possible German elections extending forward in time, one can estimate that in $1/(n+1) = 1/34 = 3\%$ of possible elections party entropy will exceed the observed maximum of 3.3 or fall below 1.7. (See Lindgren 1962:275 for the derivation, which is independent of the population distribution.) One can interpret these estimates as a measure of the extremes of German voting behavior and, in particular, an estimate of the upper limit, which is related to cognitive functioning. One can also estimate that sometime in the next hundred years or so the German party system is likely to return to a degree of fragmentation not seen since the Weimar era. This is a simple example of how one might examine the likelihood of extreme events in a political party system, such as an extreme fragmentation, which would make parliamentary government difficult.

Figure 20. Germany, 1871-1987, party entropy in relation to number of parties with parabolic curve fitting.



Notes: N = 33; R squared = 0.82; parabolic maximum estimated at 17 parties.

Figure 21. France, 1876-1978, party entropy in relation to number of parties; parabolic curve fit.



Notes: N = 30; R squared = 0.93; parabolic maximum estimated at 11 parties.

Conclusion

Reviews of past statistical research warn us to be careful about drawing theoretical inferences from observational data. Even a meticulous use of regression analysis that meets all the preconditions can lead one far astray. In this paper I have tried to give some concrete examples of how one can make interesting, definitive, and testable predictions in the area of voting behavior. Such predictions do not free us from statistical analysis but may lessen the risk of false positive results, at the same time leading to better theory construction.

I suspect that the primary critique of the examples I've presented here is that they are peripheral to concerns of many political scientists. What about predicting which party will win an election, isn't that a concern? Or political attitudes? Or the future of democracy? And so on, to any of the hundreds of topics one hears at a political science conference. I agree with that assessment.

The problem, as I see it, is that political science does not have a clear line between an identifiable core science of the discipline and politics. What is missing is the distinction that, by analogy, on finds between physics and engineering, economics and business, psychological science and psychological therapy, or physiology and medicine. True, a distinction is made between political science and public administration or, perhaps, public policy, among some academic communities. But political science tries to encompass a range of phenomena and concerns that go far beyond what any scientific theory will ever explain. This leads to over-expectations for a science of politics and frustration when the methods we use do not produce the hoped for results. Often one can and should apply scientific methods to an applied problem, but that does not make it a scientific concern. I believe we will be more productive as scientists if political science more carefully defines its scope as a science and does not try to impose unrealistic standards outside that range.

References

Ahn, Woo-kyong, Charles W. Kalish, Douglas L. Medin, and Susan A. Gelman. 1995. "The role of covariation versus mechanism information in causal attribution." *Cognition* 54: 299-352.

Blais, Andre. 2000. *To Vote or Not to Vote: The Merits and Limits of Rational Choice Theory.* Pittsburgh, PA: University of Pittsburgh Press.

Buehner, Marc J., Patricia W. Cheng, and Deborah Clifford. 2003. "From covariation to causation: a test of the assumption of causal power." *Journal of Experimental Psychology: Learning, Memory and Cognition* 29:1119-1140.

Chatfield, C. 1995. "Model uncertainty, data mining and statistical inference." *Journal of the Royal Statistical Society Series A* 158: 419-466.

Cheng, Patricia W. 1997. "From covariation to causation: a causal power theory." *Psychological Review* 104: 367-405.

Cialdini, Robert B., Raymond R. Reno, and Carl A. Kallgren. 1990. "A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places." *Journal of Personality and Social Psychology* 58: 1015-26.

Clauset, Aaron and Maxwell Young. 2005. "Scale invariance in global terrorism." http://arxiv.org/abs/physics/0502014.

Coleman, Stephen. 2004. "The effect of social conformity on collective voting behavior," *Political Analysis* 12: 76-96. For ancillary materials see http://polmeth.wustl.edu/polanalysis/ancillary12.html.

----. 2002. "A test for the effect of conformity on crime rates using voter turnout." *The Sociological Quarterly* 43: 257-276..

----. 1995. "Dynamics in the fragmentation of political party systems," *Quality and Quantity* 29: 141-155.

----. 1993. "Cycles and chaos in political party voting." *Journal of Mathematical Sociology* 18 (1993): 47-64.

----. 1985. "The human brain, social conformity, and presidential elections." *Journal of Mathematical Sociology* 11: 95-130.

----. 1975. *Measurement and Analysis of Political Systems: A Science of Social Behavior*. New York: John Wiley & Sons.

Coles, Stuart G. 2001. *An Introduction to Statistical Modeling of Extreme Values*. New York: Springer Verlag.

Dharmapala, Dhammika and Michael McAleer. 1996. "Econometric methodology and the philosophy of science." *Journal of Statistical Planning and Inference* 49: 9-37.

Hauser, J. R. and B. Wernerfelt. 1990. "An evaluation cost model of consideration sets." *Journal of Consumer Research* 16: 393-408.

Heckman, James. 2000. "Causal parameters and policy analysis in economics: a twentieth century retrospective." *Quarterly Journal of Economics* 115: 45-

Ioannidis, John P.A. 2005. "Contradicted and initially stronger effects in highly cited clinical research." *Journal of the American Medical Association* 294: 218-228.

Iyengar, Sheena S. and Mark R. Lepper. 1999. "Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology* 76: 349-366.

Kirchgaessner, Gebhard and Anne Meyer zu Himmern. 1997. "Expected closeness and turnout: An empirical analysis of German general elections, 1983-1994." *Public Choice* 91: 3-25.

Knack, Stephen and Martha E. Kropf. 1998. "For shame! The effect of community cooperative context on the probability of voting." *Political Psychology* 19: 585-99.

Knack, Stephen. 1992. "Civic norms, social sanctions, and voter turnout." *Rationality and Society* 4: 133-56.

Lindgren, Bernard W. 1962. *Statistical Theory*. New York: Macmillan.

Longford, N. T. 2005. "Editorial: Model selection and efficiency—is 'Which model . . . ?' the right question?" *Journal of the Royal Statistical Society Series A* 168: 469-472.

Mackie, T. T. and R. Rose. 1991. *The International Almanac of Electoral History* (3rd ed.). Washington, D.C.: Congressional Quarterly.

Malinetskii, G. G. and S. P. Kurdyumov. 2001. "Nonlinear dynamics and the problems of prediction." *Herald of the Russian Academy of Sciences* 71(2): 94-113. Retrieved on 7 April 2005 from http://www.keldysh.ru/departments/dpt_17/eng/ndeng.htm.

Manson, JoAnn E. et al. 2003. "Estrogen plus progestin and the risk of coronary heart disease." *New England Journal of Medicine* 349: 523-534.

Miller, George. 1956. "The magical number seven, plus or minus two: some limits on our capacity for processing information." *Psychological Review* 63: 81-97.

Miller, Henry and James Bieri. 1963. "An informational analysis of clinical judgment." *Journal of Abnormal and Social Psychology* 67: 317-325.

Pesaran, M. Hashem and Ron Smith. 1995. "The role of theory in econometrics." *Journal of Econometrics* 67: 61-79.

Plous, Scott. 1993. *The Psychology of Judgment and Decision Making*. New York: McGraw-Hill.

Rimm, Eric B., Meir J. Stampfer, Alberto Ascherio, Edward Giovannucci, Graham A. Colditz, and Walter C. Willett. 1993. "Vitamin E consumption and the risk of coronary heart disease in men." *New England Journal of Medicine* 328: 1450-1456.

Smith, Vernon. 2003. "Constructivist and ecological rationality in economics." *American Economic Review* 93(3): 465-508.

Sornette, Didier. 2002. *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton NJ: Princeton University Press.

Summers, Lawrence H. 1991. "The scientific illusion in empirical macroeconomics." *Scandinavian Journal of Economics* 93(2): 129-148.

Vivekananthan, D.P. et al. 2003. "Use of antioxidant vitamins for the prevention of cardiovascular disease: meta-analysis of randomized trials." *Lancet* 361:2017-2023.

July 2005