

Towards an Information Entropy Model of Job Approval Rating: The Clinton Presidency

Rajendra G. Kulkarni^{1*}, Roger R. Stough² and Kingsley E. Haynes³

¹ Corresponding Author, Research Instructor, Mason Enterprise Center, The Institute of Public Policy, George Mason University, Fairfax, VA 22030, USA

Tel.: 703-993-3177, Fax: 703-993-2284, E-mail: rkulkarn@gmu.edu

² NOVA Endowed Chair, Eminent Scholar and Professor of Public Policy, Director, Mason Enterprise Center, The Institute of Public Policy, George Mason University, Fairfax, VA 22030, USA

³ Director, The Institute of Public Policy, University Professor/Eminent Scholar, Professor of Decision Sciences, Geography and Public Affairs, George Mason University, Fairfax, VA 22030, USA

*Author to whom correspondence should be addressed.

Received: 3 September 1999 / Accepted: 27 September 1999 / Published: 30 September 1999

Abstract: This paper discusses an analytical approach to explaining a nearly constant high job approval rating of president Clinton between January, 1998 and February, 1999. Despite all the controversy and massive information exposure to mostly unflattering news about Mr. Clinton; the public, in nearly all major opinion polls expressed their wish that Mr. Clinton be allowed to complete his second term in the office. The analytical approach is based on the information entropy theory of Shannon. The model is tested using the data from the polling archives of ABC/Washington Post. The results are confirmed by the Kendall's τ statistics.

Keywords: Clinton Presidency, President's job approval rating, Public opinion polls, Information entropy, Relative entropy

1. Introduction

Free societies attach great importance to public opinion, and rightly so. Its *legitimacy* is beyond doubt. The role played by public opinion in the conduct of a modern free state is well known [1]. One way to find a free society's expression is through opinion polls that are useful in understanding the public's attitude towards specific issues at a particular point in time.

The rise of mass-media after the 2nd World War made public opinion polls almost impossible to discount. Compared to other democratic societies the public participation in policy formulation has been relatively higher from the time the U.S. was a young nation. In the early years after the birth of this nation, the Jeffersonian view of a bottom-up approach had an edge over the top-down decision making propounded by both Hamilton and Madison. One of the plausible explanations for this edge is the American Revolutionary War against the British monarchy, which was a rejection of that symbol of authoritarian control of the affairs of the state. Both the French statesman Tourquville (in the early 18th century) and the British philosopher Bryce (in the late 18th Century) have widely commented on the U.S. public participation in policy making. Over time mass education contributed even more to the notion of participatory citizenry in the governance and public affairs at all levels of society. However, the contribution of public opinion to policy formulation was more theoretical than empirical. The middle of 20th Century saw the emergence of the new discipline in of Political Science, armed with empirical and analytical tools and mass-media based polling techniques as a direct result of rise of mass-media after the 2nd World War. For the first time, polling techniques could be used to measure the public's views on policy issues, and ever since then public opinion polls have become a major player in the states' affairs. However, between the theory of public participation in governance (pre 2nd world war) and practical tools for measuring its contribution is the almost hidden process of how mass opinions are formed and how consensus (or lack of it) among the public builds over various issues [2]. Further, once the public reaches a consensus what is the process by which such a consensus is maintained?

2. Background

In this paper we analyze the public's high job approval rating for President Clinton. Nearly every major public opinion poll since January 1998 has given the President a job approval rating of near 60% and above (Figure 1)

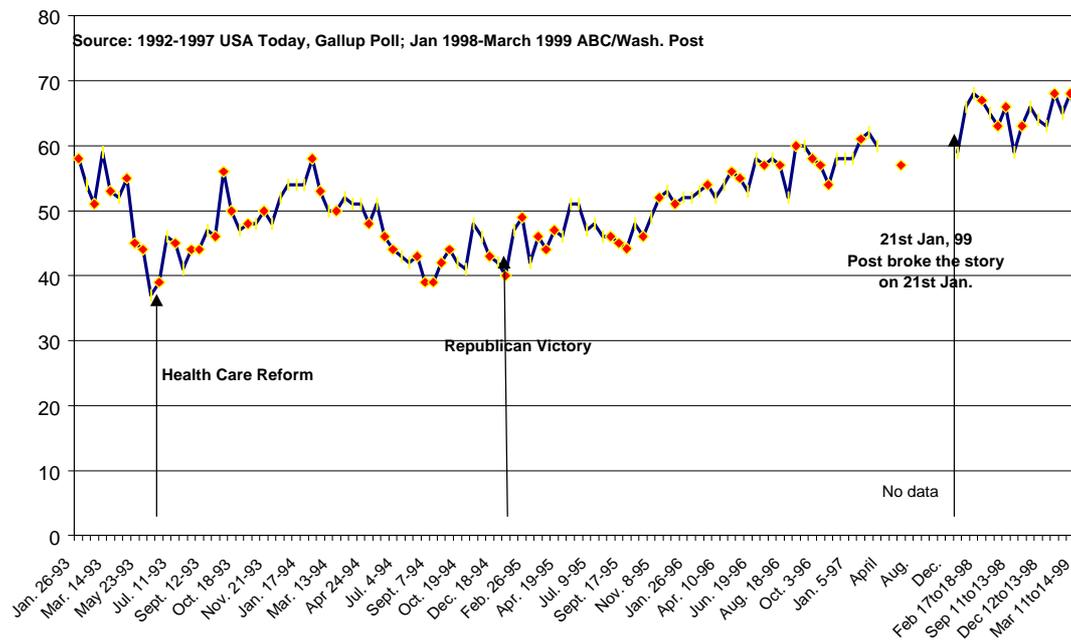


Figure 1. President's job approval rating by selected dates: 1992 to 1999.

Such high ratings are unusual for the following reasons.

- Historically, second term presidents lose their popularity especially when the presidency is already being branded by the elite as a "lame duck."
- If a presidency has been dogged by one or another scandal from the day of inauguration (the Whitewater, the travel office, FBI files, the presidential campaign fund controversy, and at least half a dozen Independent prosecutor investigations against cabinet secretaries). These begin to take a toll on popularity as individual power and election support attenuates.
- American politics do not allow second acts [3], i.e., if you make a mistake you often pay the price with your job or at least your popularity.
- Almost no one can retain his/her approval ratings in the face of an unfolding and hence continuous scandal (Watergate: Nixon's approval rating fell to 31% in August, 1973, a 36% drop in six months; Irangate: during the Carter presidency, 31% by October, 1979; Iran-Contra during the Reagan presidency: job approval fell to 43%, a 20% drop by March, 1987; [4].)
- Unlike other scandals the Lewinsky scandal had the recency effect, in the sense it was something that did not happen a long time ago.
- Unlike other scandals where a president's involvement was at the most circumstantial and peripheral the Lewinsky scandal offered no such defense.
- It was impossible to ignore the scandal because there was no escape from the multimedia (TV, newspapers, magazines and internet) exposure and round the clock coverage, most of which was not flattering to the president.

Several explanations have been offered to explain the President's continuing high approval rate. Much has been made of the correlation between good economic times and high approval rating of a sitting president. In fact this has been used as the main explanatory variable for the president's popularity, i.e., people feel good about their economic well being and give credit to the President. Some have talked about scandal weariness among the people. Still others have said that America was becoming more like Europe where personal indiscretions do not affect the public's view of politicians. A variation of that has been that, as long as there is no threat to national security people do not mind the not-so-perfect private lives of their elected officials. Yet others have offered the explanation of how the partisan tone of the scandal and the seemingly unfair treatment of the president by the opposition made people sympathetic towards the president. Some have talked about cultural and moral decline in the society. In short there have been many explanations, possibly each of them has played a part in giving the president such high job approval. So what is driving such high poll numbers? A model based on Shannon's information entropy is presented in this paper. The results are tested by the data from the ABC/Washington Post's polling archives [5] and its validity is confirmed using Kendall's τ statistics [6].

Opinion polls are snap shots of the public's perception. As stated earlier, every major public opinion poll between January 1998 and February 1999 gave the president a very high job approval rating, reflecting the public's perception of his ability to do the work. It was not as if public did not know about the scandal. One could not have escaped the media coverage of the scandal. So what is the explanation for giving the president such high approval? Could it be that the public filtered out all the noise that passed for minute details about the affair and concentrated mainly on the essence of the story? In other words, once the public realized in the first few days following the January 21st news, that there were no national security issues, nor any financial mismanagement, the entire story became a tragedy of human failings. Throughout the following year the basic story remained same. Every new and so-called breaking news did not seem to add any new *information* to the public's perception of the president's behavior.

As far back as the 1992 election, the plurality of the public had ignored the *character* issue and voted Clinton into the office in the three-way race. In the following years, as the public became more aware of many of his private peccadilloes it continued to separate the president's public life from his personal problems. For example, even after the reinstatement of the Jones lawsuit in May of 1996, the public seemed to have *compartmentalized* Clinton's private and public lives [7].

The following section will introduce the main points of information entropy. Section IV tests the model using archived data from public opinion polls and the results validated using the Kendall's τ statistic. Section V provides the conclusion and suggests the future research.

3. Surprise and Information Entropy

In communication theory, the information content of an event is measured in terms of surprise or the unexpectedness of the occurrence of such an event [8]. Suppose, a weather forecast for the North-East U.S. is for light showers on a spring day, and indeed it does rain. Such a forecast has little or no

information value, since the occurrence of showers during the spring season is expected or is a norm. Instead if heavy snow is forecast for spring and actually occurs, then such a forecast has more surprise and hence high information content than the forecast of spring showers. Thus, the information content of a statement is decided by the degree of *surprise* or *unexpectedness* of an event. If an event can occur in n different ways, then the amount of information obtained is higher when the outcome of the event is least expected. In other words, the least probable event has the most information. Next, let us quantify the above explanation in mathematical terms with the help of an example.

Suppose a card is drawn from a randomly shuffled deck of 52 cards then:

1. Event A refers to the drawn card being a spade and the probability of obtaining a card of spade is $p(A) = 13/52 = 1/4$. Let $I(A)$ denote the information associated with event A .

2. Event B refers to the drawn card being an ace, and the probability of obtaining an ace is $p(B) = 4/52 = 1/13$. Let $I(B)$ refer to the information associated with the event B .

3. Event C refers to the drawn card being an ace of spade and the probability of obtaining an ace of spade is given by the joint probabilities of events A and B and is

$p(A \cap B) = (1/4) * (1/13)$. And the information associated with the joint events is given by $I(C) = I(A \cap B)$.

$$p(C) = p(A \cap B) = p(A)p(B). \quad (1)$$

Obviously, the probability of drawing an ace of spades from a perfectly shuffled deck of cards is much less than the probability of drawing any ace, which is less than that of drawing a spades, i.e., $p(C) \leq p(B) \leq p(A)$. In other words,

$$I(C) = I(A \cap B) \geq I(B) \geq I(A). \quad (2)$$

Moreover the event of drawing the ace of spades has more surprise (information) than the event of drawing any ace, which in turn has more surprise (information) than drawing a spades from a deck of cards,

$$1 \geq p(E) \geq 0. \quad (3)$$

Note that the probability of any event E taking place has the following relations:

$$I(E) \geq 0 \quad (4)$$

And the information associated with that event E can never be negative, i.e.,

So what could be the function that will satisfy relations expressed in equations (1) through (4)? According to Shannon's Information theory [9], the log function will satisfy all of the above relations. Thus for an event E , a relation of between information $I(E)$ and probability of event E exists:

$$I(E) \equiv -\log_b(p(E)), \quad (5)$$

where, the constant $b > 0$ is called the base of logarithm. The negative sign helps preserve the non-negativity property of information (equation (4)). Since probability of an event is bounded between 0 and 1 (equation (3)), the logarithm function will hold both for an event that is certain and for an event that is impossible. When an event is certain (its probability is 1), the information associated with it is zero. On the other hand, when an event is impossible (its probability is zero), the associated information is infinite. In the case of an event of drawing an ace of spade the log of joint probability translates to addition of information of two events, the drawing a spade and the drawing an ace. The following equations illustrate the information relations for the above example.

$$I(A) = -\log(p(A)) = -\log(1/4) \quad (6)$$

$$I(B) = -\log(p(B)) = -\log(1/13) \quad (7)$$

$$I(C) = I(A \cap B) = -\log(p(A)p(B)) = -[\log(p(A)) + \log(p(B))] \quad (8)$$

$$\therefore I(C) = I(A) + I(B) \quad (9)$$

The example above satisfies both the logical and algebraic requirements in the computation of information content.

Next let us consider an event that has n number of possible outcomes defined by a probability distribution function $f(p_i)$ over all $i=[1, n]$. Then what is the average or expected information H of all of these events? H is computed as follows:

$$H(p_1, p_2, \dots, p_n) = -\sum_i^n p_i \log(p_i) \quad (10)$$

Equation (10) gives the mean or average information associated with an event that is defined in terms of its probability function $f(P)$. It is also known as the *information entropy* of the event, an expression of the degree of randomness in the occurrence of an event. If one does not have an *a priori* distribution of the probabilities of an event then using the principle of symmetry [10], one must assume that the event occurs with uniform distribution of probability [11]. Such probability distributions are referred to as *non-informative* because all outcomes are equally possible, and such distributions offer minimum information or maximum entropy. For example if an event K can occur in n ways then the probability of each of these outcomes is $1/n$ and hence the information entropy H_{max} is given by:

$$H_{max} = - \sum_1^n (1/n) \log(1/n) = -\log(1/n) = \log(n) \quad (11)$$

Equation (11) gives minimum average information or maximum information entropy of the event K . Suppose, that there is a way to assign a probability distribution because we have information about how the event K occurs, i.e., we have an alternate probability distribution of event K then the let information entropy associated with such distribution be, H_c , then the relative entropy of information H_R is computed as follows:

$$H_R = H_c / H_{max} \quad (12)$$

and the organization S is a measure of order or lack of disorder defined as $1 - H_R$ [12 Haynes]. Such order is introduced due to *knowing* an alternate probability distribution of event K as opposed to uniform probability distribution and is computed as follows:

$$S = 1 - H_R. \quad (13)$$

An example due to Jessop [8] will make these ideas a little more clear. The Maximum entropy associated with the English language with 26 letters and an average of 8 letters per word is 3.296. This would be equivalent to having a language where all alphabets were equally possible in all 8 letter words, a language with no patterns in the letters and the words. However, the English language does exhibit a pattern which makes it intelligible and hence the structured nature of the English language can be expressed in terms of relative entropy as $(2.852/3.296) = 0.865$. Using equation (13) one may obtain a proxy for computing the organization of the English language as $1 - 0.865 = 0.135$.

In the next section let us apply the concepts of information entropy to the public opinion polls conducted over a one year period across the nation and its regions, namely the Eastern, the Midwestern, the Southern and the Western. Using the definition of surprise or unexpectedness, the concept of maximum entropy, the relative entropy and the order, we can analyze the constancy of the job approval rating of Mr. Clinton between January 1998 and March 1999 [5].

4. Application of Information Entropy to the Polling Data

Suppose, that as part of public opinion polls an unbiased sample of 100 people are selected and each one of them is asked to record his/her opinion to the following question:

Do you approve or disapprove of the way the president is doing his job? Of the three select one:
Yes, No, Don't know.

Before conducting such a poll, the outcomes according to the principle of symmetry can be: $1/3^{\text{rd}}$ of the respondents approve; $1/3^{\text{rd}}$ do not approve and $1/3^{\text{rd}}$ say they don't know. Then the *a priori* amount of information entropy or disorder is given by:

$$H_b = -\sum (1/3)\log(1/3) = \log(3) = 0.4771 \quad (14)$$

But as soon as the public opinion is known then the amount of information entropy or ignorance is subject to the actual number of people that answered yes, no and don't know. For example, in the 1992 election, Mr. Clinton got 43% of the votes, Mr. Bush got 38% and Mr. Perot got 18%. Then using the vote percentages as the proxy for expressing the approval of a candidate, the information entropy for 1992 is:

$$H_{1992} = -[0.43 \times \log(0.43) + 0.38 \times \log(0.38) + 0.19 \times \log(0.19)] = 0.4543 \quad (15)$$

A little improvement is found in equation (15) over the unbiased ignorance of equation (14), since the public *knows* a little more about the candidate Clinton. In the four years since 1992, the public became more informed about Mr. Clinton and the poll numbers for the 1996 presidential elections reflect the public's approval/disapproval. The following is the breakdown of the percentage of votes received by each candidate: 49% for Mr. Clinton; 41% for Mr. Dole 41% and 9% for Mr. Perot. In other words, using the same question from the public opinion poll above, these election votes can be seen as 49% saying Yes, 41% saying No and 9% saying Don't Know. Using this data, we obtain from equation (11),

$$H_{1996} = -[0.50 \times \log(0.50) + 0.41 \times \log(0.41) + 0.09 \times \log(0.09)] = 0.4034 \quad (16)$$

Thus by 1996, the information entropy or the lack of ignorance was already down by about ~15%. After the election and all through 1996 Mr. Clinton continued to get better job approval numbers than ever before (See Figure 1, based on data [5] and [7]). By the end of 1997, 59% approved of the job he was doing, the equivalent of information entropy H_{1997} of 0.3508, or the ignorance of his handling of his job was down by about 14% from the 1996 election or down by about 27% from the 1992 election. Between 1992 and 1997, the Clinton Administration was dogged by one or another kind of scandal. In spite of such bad publicity, by the end of 1997 enough people knew about Clinton to compartmentalize his public and private lives. In fact, by normalizing the current entropy level with respect to the 1992 information entropy (H_{1992}), the relative entropies of 1996 and 1997 can be computed using equation (12) as:

$$H_{1996} = 0.85 \text{ and } H_{1997} = 0.73 \text{ respectively.}$$

Further, the measure of order or organization in Mr. Clinton's job approval rating by the end of 1997 can be computed from equation (13) and is found to be nearly 0.2647.

5. The Entropy of Polling Data between January, 1998 and February, 1999

Figure 2 shows Mr. Clinton's job approval rating for selected dates between 1998 and 1999 (data from [5]).

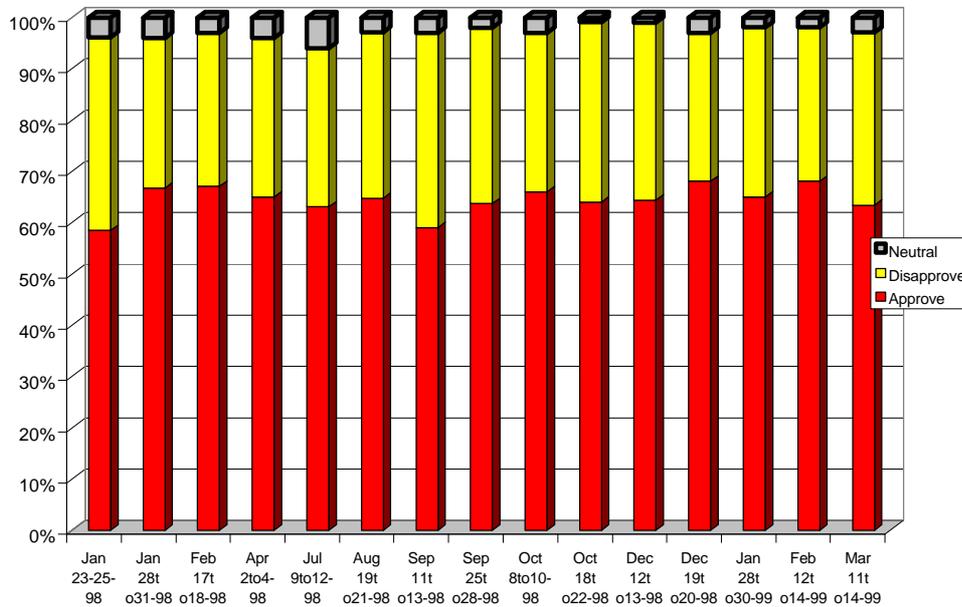


Figure 2. President`s job approval/disapproval rating by selected dates.

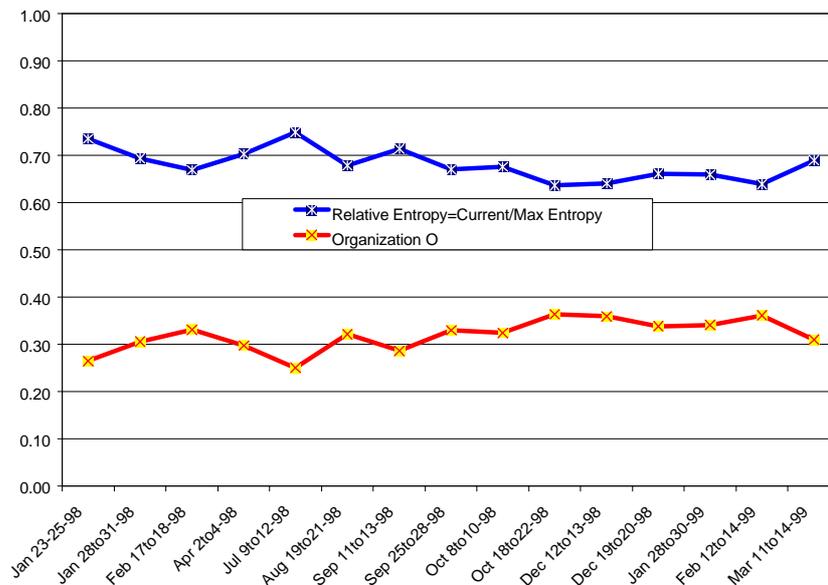


Figure 3. Relative info. entropy and order S of president`s job approval rating by selected dates.

The same data has been used to compute the relative entropy R_H and the order S . These two quantities are complementary (Figure 3). An R_H of 1 indicates complete lack of order S among the public.

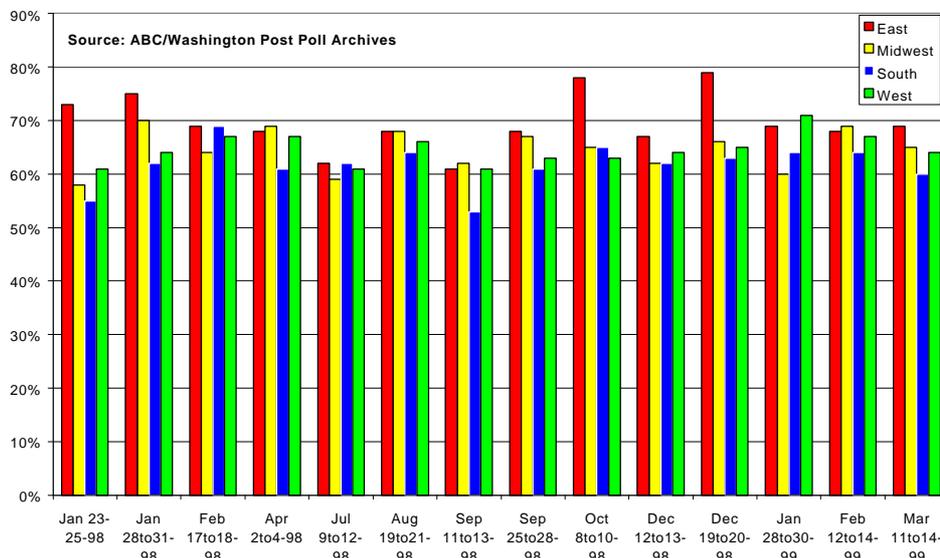


Figure 4. President's job approval by region by selected dates.

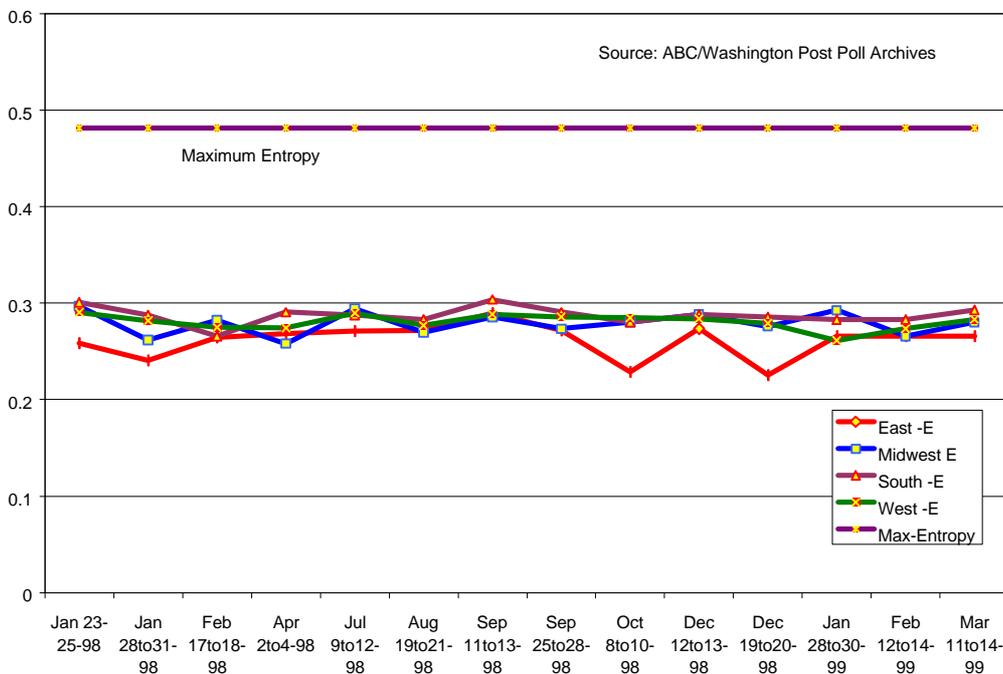


Figure 5. Regional entropy of president's job approval/disapproval rating.

Thus even an average $R_H \sim 0.7$ with a variance of 0.0011 indicates that the public though not comfortable with the conduct, did not want Mr. Clinton to be removed from the office. Figure 4 shows the regional approval data and Figure 5 shows the relative magnitudes of the maximum entropy and the regional entropies. The East region has the least amount of entropy. The South, the Midwest and the West regions all converge towards 0.3 and below entropy levels. Similarly, near constant figure are computed for the polling data by education and by sex. Even the party affiliations data shows a similar trend, as if there was a lock in among the Democrats and the Independents in favor of the President, along with a tiny number of Republicans, while nearly 60% of Republicans disapproved of Clinton's job performance.

Validation of the data using the Kendall's τ statistics

Kendall's τ statistics [6] measures trends in the time series data that has been collected a periodically. If the computed statistic is zero or near zero then there is no trend, suggesting a near constancy of the values in the time series. If the statistics tends to be near +1 then there is an upward or increasing trend in the time series values and a decreasing or downward trend of the statistic tends to be near -1. Using data from Figure 1, it was found that the statistic remained nearly zero (a bit negative) with a value of -0.04762 between January of 1998 and February of 1999, in agreement with the near constant value of the information entropy and near constant organization factor (See Figure 2). Between January of 1992 and December of 1994, the τ statistic shows a negative value of -0.14296, a downward trend and between January of 1995 and December of 1997 the τ statistic shows a value of +0.33913, an upward trend. But after the breaking of the story in January 1998 and onwards the statistics stays steady with near zero value of -0.04762. All of these results agree with the charts shown in Figures 1 and 2. Similar statistics were computed for the regional data, both for approval and disapproval numbers. The tables 1 and 2 below show the computed values for approval and disapproval respectively by the regions.

Table 1. Approval τ statistic.

Approval	East	Midwest	South	West
Kendall's τ	-0.12	-0.19	0.05	0.08

Table 2. Disapproval τ statistic.

Disapproval	East	Midwest	South	West
Kendall's τ	-0.16	0.05	-0.16	-0.30

Both the figures in the tables show a lock in of the statistics; only the West has a slightly higher negative value, which suggests that the disapproval numbers decreased over the time period. The data in the tables confirms the regional constancy of the information entropy, as shown in Figure 4.

6. Conclusions and Future Research

The information entropy model offers a way to test the public's confidence level in a public figure despite the controversy surrounding the issues. In democratic societies a majority vote is accepted as a sign of victory. But in reality, even a super majority may not settle the issues completely. For example, even a three way vote distribution of 75%, 15% and 10% results in an entropy $H = 0.31$, or $H_r = 0.67$, a reduction of just 33% from the maximum entropy. It may be interesting to investigate what level of entropy translates into total acceptance of an issue by the public. The model needs to be tested on other situations and types of data where there is a vast of number of choices expressed by the public and adding values or weighs to different opinions. Additionally, the model needs to be augmented with an opinion formation model. An attempt will be made to integrate the current model with the one developed by Paelinck [13]. The concept of organization factor S needs to be developed further and tested. A link between organization factor S and self-organizing systems will be explored.

Acknowledgement: The authors wish to express their sincere thanks to Dr. J.H.P. Paelinck (J.Paelinck@poboxes.com) and Ms. Mary Clark (mclark7@gmu.edu) for providing timely advice and guidance for this paper.

References and Notes

1. Yelic, J. L.; Todd, J.R. *Public Opinion: the visible politics*, F.E. Peacock publishers: Itasca, Il., 1989.
2. Yankelovich, D. *Coming to Public Judgement: making democracy work in a complex world.*, Syracuse University Press: Syracuse, N.Y. 1991.
3. This adage is a variation of a quote from Scott Fitzgeralds: "There is no second act in American lives."
4. (a) Gallup Organization, 1999. <http://www.gallup.com/poll/release/pr980211.asp> (b) Dr. George Gallup pioneered public evaluation of presidents with President Roosevelt in the 1930s. Since then, the basic question, "Do you approve or disapprove of the way [name of president] is handling his job as president?" has become the standard in political polling.
5. ABC/Washington Post Polling Archive:
<http://www.washingtonpost.com/wp-srv/politics/polls/polls.htm>
6. (a) Kendall, M.G. *Rank Correlation Methods*, Hafner Publishing Company, Inc.: NY, 1962. (b) Gibbons, J. D. *Nonparametric Statistical Inference*, McGraw-Hill Book Company: NY, 1971.
7. USA Today: <http://www.usatoday.com/elect/eq/eqindex.htm>
8. (a) Jessop, A. *Informed Assessments* , Ellis Harwood: New York, NY. 1995. (b) Haynes, K.E.; Phillips, F. Y.; Mohrfield, J.W. The Entropies: Some roots of Ambiguity. *Socio-Economic Planning Structures* **1980**, *14*, 137-145.

9. Shannon, C.; Weaver, W. *The Mathematical Theory of Communication*, Univ. of Illinois Press: Urbana and Chicago, IL. 1963.
10. Jaynes, E.T. Where do we stand on maximum entropy in *The Maximum Entropy Formalism*. Eds: R.D. Levine and M. Tribus. MIT Press: Cambridge, MA, 1979; pp15-118.
11. Applebaum, D. *Probability and Information: An Integrated Approach*, Cambridge Univ. Press: Cambridge, U.K. 1996.
12. Hyanes, K. E.; Storberk, V. The Entropy Paradox and the Distribution of Urban Population. *Socio-Economic Planning Structures* **1978**, *12*, 1-6.
13. Paelinck, J. H. P. Working paper, *Model of public opinion: A Fuzzy Probability Approach to Opinion Diffusion Dynamics*. JHPP/WPS2/WP6/02.15.99.

© 1999 by the authors. Reproduction of this article, by any means, is permitted for noncommercial purposes.