Whose Ideas? Whose Words? Authorship of Ronald Reagan's Radio Addresses

Edoardo M. Airoldi, *Princeton University*
Stephen E. Fienberg, *Carnegie Mellon University*
Kiron K. Skinner, *Carnegie Mellon University and Hoover Institution*

> If it was not possible to arrive at our present understanding of Eisenhower's leadership until the archives were open, does it follow that we are likely some day to find evidence documenting the existence of a Ronald Reagan whose political style was drastically different from what it now appears to have been?
>
> —Fred Greenstein (1990)

On January 5, 1975, Ronald Reagan completed two terms as governor of California, and soon thereafter he began taping his nationally syndicated radio program. Between January 1975 and October 1979, with the exception of a brief interlude to compete for the Republican presidential nomination in 1976, the former governor delivered three-minute commentaries that were broadcast every week by approximately 300 U.S. radio stations, reaching an audience of between 20–30 million listeners.

When a collection of Reagan's handwritten radio commentaries was published on his 90th birthday (February 6, 2001), a firestorm ensued. Many scholars, statesmen, and policy analysts who read *Reagan, In His Own Hand* (Skinner, Anderson, and Anderson 2001)[1] were astonished to learn that Reagan had been thinking, speaking, and writing about the major domestic and foreign policy issues facing the United States long before he launched his presidential campaign in 1980.[2] The newly published documents revealed that years before he took presidential office, Reagan had crafted detailed policy positions and outlined his economic and political philosophies. A further revelation was the degree to which Reagan's performance as president, including his policies toward the Soviet Union, reflected the messages contained in these earlier radio broadcasts.

Ronald Reagan, the 40th president of the United States, may well be remembered best for his contribution to ending the Cold War: "Commonly seen as a prototypical conservative, Reagan proved capable of truly radical thinking when it came to nuclear weapons, as evidenced by his apparently sincere desire to abolish nuclear arms and replace deterrence with strategic defense" (George and George 1998). What is perhaps most interesting about Reagan's Soviet policy is that it was at odds with the accepted wisdom about deterrence.

The U.S. strategy of deterrence during the Cold War consisted of placing strict limits on the research, development, and deployment of missile defense systems and Mutual Assured Destruction (MAD), the doctrine designed to prevent nuclear war by allowing the populations of both the U.S. and the Soviet Union to be vulnerable to a nuclear attack. On March 23, 1983, President Reagan announced that he was authorizing a research and development program known as the strategic defense initiative (SDI). This program was Reagan's explicit attempt to begin breaking out of MAD and the limits on missile defense outlined in the Anti-Ballistic Missile (ABM) Treaty of 1972. The source of Reagan's anti-classical thinking about nuclear policy is of interest to policy elites (see, for example, FitzGerald 2000).

In the radio commentaries he wrote several years before he became president, Reagan criticized MAD and the ABM Treaty. During his March 23, 1977, broadcast, he remarked sarcastically that "one of . . . [the United States'] contributions to détente" included bargaining away the right to develop missile defense, as specified by the ABM Treaty. The following year, in his commentary of June 27, 1978, he spoke of the Soviets' refusal to "subscribe to our belief in 'mutual assured destruction'" (Skinner, Anderson, and Anderson 2001, 79, 119). The statistical analysis reported here takes us closer to verifying Reagan's claim, made in a letter dated July 7, 1988, that SDI "was my idea to begin with" (Skinner, Anderson, and Anderson 2001, 431).

Kiron Skinner, an editor of five volumes of Reagan's writings and a co-author of this article, found approximately 679 handwritten drafts of the more than 1,000 radio commentaries Reagan delivered in the late 1970s. It is known that some of the typed radio essays were composed by aides, such as Peter Hannaford, but the absence of handwritten drafts made it difficult to determine whether Reagan was the author of most of the other 300+ commentaries. We undertook a statistical analysis to determine the authorship of these remaining radio addresses (Airoldi, Anderson, Fienberg, and Skinner 2006).

The analysis that follows is a semi-technical review of our statistical research and our results using word counts of the radio speeches. We worked with an electronic version of the Reagan speeches and, through the availability of handwritten drafts, were able at the outset to attribute 679 speeches to Ronald Reagan and 39 to his collaborators (12 to Peter Hannaford, 26 to John McClaughry, and one to Martin Anderson). Authorship of the remaining 312 speeches is uncertain. As Reagan's main collaborator on the radio addresses, Hannaford probably wrote the initial drafts of those Reagan

Edoardo M. Airoldi *is a postdoctoral fellow at Princeton University, affiliated with the Lewis-Sigler Institute for Integrative Genomics, and the department of computer science. His research interests include statistical methodology and random graph dynamics, with application to the social and biological sciences.*

Stephen E. Fienberg *is Maurice Falk University Professor of Statistics and Social Science in the department of statistics, the machine learning department, and Cylab, all at Carnegie Mellon University. His research focuses on the development and application of statistical methodology, especially for problems involving categorical data.*

Kiron K. Skinner *is associate professor of history and political science at Carnegie Mellon University and research fellow at the Hoover Institution, Stanford University. She serves on the Chief of Naval Operations Executive Panel and the National Security Education Board.*

 DOI: 10.1017/S1049096507070874

https://doi.org/10.1017/S1049096507070874 Published online by Cambridge University Press

Table 1
Authorship of Reagan's Radio Addresses, 1975–1979. Breakdown of the Available Texts by Author and Year

Author	1975	1976	1977	1978	1979	Total
R. Reagan	60	195	52	219	153	679
P. Hannaford (radio)	1	5	2	4	0	12
P. Hannaford (news)	5	0	7	18	0	30
J. McClaughry	0	3	1	15	7	26
M. Anderson	0	0	0	0	1	1
Total author uncertain	149	80	4	25	56	314
Total known author	66	203	62	256	161	748
Total (all)	215	283	66	281	217	1,062

did not write, although a few other people contributed possible drafts. Because of this uncertainty, we also coded several of Reagan's newspaper columns known to have been drafted by Peter Hannaford (see Table 1).

Methodology

Over the last decade, the analysis of textual data has been a central research theme in the computer science community. New methodology has appeared, and existing models and algorithms which were successful in analyzing other types of data have been adapted to analyze text. In spite of the extreme diversity in approaches, they all share a common representation of electronic documents in the basic units of analysis, usually described as "bags of words." According to such representation, documents become collections of word counts "drawn from the bag" of the size of a known vocabulary. Therefore, they are amenable to statistical modeling. Depending on the context, some applications have been more successful than others. A popular example is that of an automated system capable of categorizing "newsgroup posts" by "topic" (Mitchell 1997, chapter 6). In the past few years, successful case studies in the political science literature have taken advantage of this simple, but powerful, representation of documents. For example, Michael Laver, Kenneth Benoit, and John Garry (2003) proposed a method for extracting policy positions of political parties from political texts; the policy positions recovered in this fashion match those estimated by expert analysts to a satisfactory degree. Another example is given by recent work of Cheryl Schonhardt-Bailey (2005), who mapped George W. Bush and John Kerry's speeches on national security to a hypothetical space defined by words and their frequencies; the position of the words chosen by presidential candidates in such a space is suggestive of substantive interpretations.

But the idea of "using statistics about the usage of words to inform the analysis of problems regarding text" has a long history that predates the interest of the computer science community by more than a century, going back to De Morgan (1872) and especially Mendenhall (1887), who set out to capture the essential elements of the writing style of an author by noting the frequency patterns of words throughout the text. Indeed, models for word counts have a long tradition in the intervening years.

While computer science approaches favor "black boxes" that lead to low prediction error, statistical approaches pay a great deal of attention to the assumptions of the model and to the interpretability of parameters and results. In both cases, the models provide a useful approximation to summarize the data and make predictions about the author of a document, for example. If the statistical assumptions hold, the prediction error tends to be low and we can attempt a convincing interpretation of the findings, as the modeling choices are meant to capture semantics of interest. For example, in their pioneering Bayesian statistical analysis of the *Federalist Papers*, Frederick Mosteller and David Wallace (1964; 1984) captured relevant aspects of the authorship attribution problem they aimed to resolve with their every modeling choice. They then pursued a careful word selection strategy that was essential to produce believable predictions.

In our study, we explored the full gamut of modern exploratory tools, and in the process learned many features that held potential promise in distinguishing speeches written by Reagan from those written by Hannaford and others. But in the end, we settled on a formal model rooted in Mosteller and Wallace's pioneering work (1964; 1984).

Our Model

Following Mosteller and Wallace, we posited a fully Bayesian model to deal with the problem of authorship attribution that relaxes some of the assumptions established in modern computational analyses of textual data, such as the fact that occurrences of a word across the text are assumed independent. In particular, we considered two specific models of word counts that are meant to capture the notion of an author's writing style: the Poisson and Negative-Binomial models. The idea is that authorship emerges through patterns of language choices. For example, as a text progresses, an author's personal writing style creates a context within which words and idioms that have already appeared are more likely than others to reappear.

What are these models for writing styles? Let X_w be the random variable that encodes the number of times word w appears in a document, with expected rate of occurrence $\Theta_w = \omega\mu$; where μ is the expected rate of occurrence of a word in a document of reference length λ, and ω is the observed length of a document as a multiple of the reference length. The reference length λ is itself a parameter of the model and we need to either estimate it or choose it carefully.[3] In this revised notation, the Negative-Binomial distribution is simply a Poisson distribution with extra variability controlled by the parameter $\delta = \mu/\kappa > 0$. We leverage this fact and make the connection explicit in our parameterization:

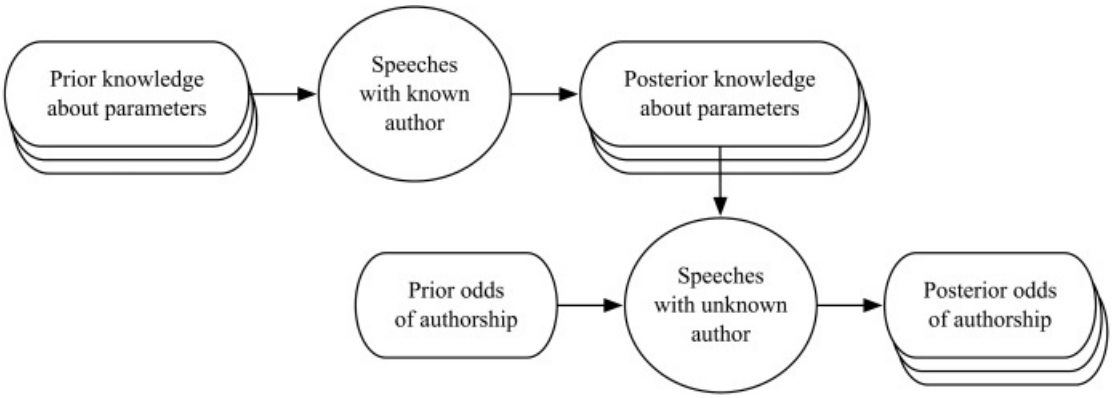
$$Poisson(X_w = x|\Theta_w = (\omega, \mu)) = \frac{e^{-\omega\mu}(\omega\mu)^x}{x!},$$

$$Negative\text{-}Binomial(X_w = x|\Theta_w = (\omega, \mu, \delta)) = \frac{\Gamma(x+k)}{x!\Gamma(k)}(\omega\delta)^x(1+\omega\delta)^{-(x+k)}.$$

In terms of (ω, μ, δ) we used for the Negative-Binomial model, δ seemed stable across words and authors—mostly $\delta \in [0,0.75]$ with some heavy tails. Such heavy tails in the extra-Poissonness parameter δ are mostly due to personal pronouns, but we included them in the analysis nonetheless because it turns out that they make good discriminators. In order to use a simple prior for δ (see below), we used a variance stabilizing transformation to reduce the heavy tails as in $\zeta = \log(1 + \delta)$. The assumption that δ is the same for all authors was satisfactory for most words but not for those that occurred with low frequency. Even though differential non-Poissonness is potentially discriminating, our actual motivation for the choice of modeling

https://doi.org/10.1017/S1049096507070874 Published online by Cambridge University Press

Figure 1
An Overview of the Strategy for Estimating Models' Parameters, and Predicting Authorship of the Uncertain Speeches[5]



possibly distinct δ parameters was to avoid upsetting the analysis.

Eventually, we introduced a different parameterization that allowed for prior information about the differential use of words by two authors to be expressed in a natural way, and for this differential information to be captured by priors with a simple functional form. In the case of two authors, the re-parameterization is as follows:

$$\sigma = \mu_1 + \mu_2 \qquad \xi = \zeta_1 + \zeta_2$$
$$\tau = \mu_1/\sigma \qquad \eta = \zeta_1/\xi.$$

The rationale behind this transformation from $(\mu_1, \mu_2, \zeta_1, \zeta_2)$ to $(\sigma, \tau, \xi, \eta)$ is that we are separating the "differential usage" of words, which is crucial for making inferences about the authorship of a document, from the "overall usage" of words, which is mostly needed to calibrate the probability models of word occurrences. Once the parameters that relate to these two tasks are separated, we can posit models for prior knowledge that are non-informative about authorship for the "differential" parameters, and posit models for prior knowledge that improve model fit for the "overall" ones. For example, if a word did not appear in the training data, a natural, non-informative assumption is that both authors may be using such a word in a similar fashion. In the new parameterization in terms of $(\sigma, \tau, \xi, \eta)$, we can center the prior for (τ, η) around (0.5, 0.5) with an equal chance of deviating toward zero or one. We further explored the extent to which such an assumption would hold for other words in a study devoted to priors.

Technical Strategy for Producing Predictions

Figure 1 illustrates the strategy we adopted to predict the author of the 314 speeches of uncertain authorship. For any combination of model and parameterization, e.g., a Negative-Binomial model specified in terms of $(\sigma, \tau, \xi, \eta)$, the strategy we adopted entails two steps: (1) starting with non-informative distributions on the parameters, the texts of the speeches whose author is known are used to derive posterior distributions on the parameters; the posterior means are then taken as point estimates of the parameters in the next step; (2) starting from equal odds of authorship, the words in each of the 314 speeches of uncertain authorship are combined with parameter estimates obtained in the previous step to produce posterior odds of authorship. The posterior odds computed for each of these speeches summarize the information that is needed to make a prediction about its authorship.[4]

In order to carry out step (1), we had to fix the values of the constants β underlying the non-informative prior distributions for the parameters. We adopted both a full Bayesian approach and an empirical Bayes approach (Carlin and Louis 2000). In the full Bayesian approach, which we prefer as it allows for an analysis of the sensitivity of the predictions, we predicted the authorship for each text for 21 plausible sets of values for the underlying constants that were selected in the study of priors. In doing so, we obtained a set of 21 predictions for the text of each speech of uncertain authorship, which we combined by means of the majority rule. The empirical Bayes approach produced only one set of estimated values for the underlying constants, and this fell roughly in the middle of the 21 sets of values from the full Bayesian approach.

Word Selection Strategies

The task of word selection was complementary to model building . As with the model, a careful strategy to select the word that best captures the differential style of Reagan versus that of his close collaborators is essential to produce believable predictions. To that goal, we aimed at finding a small set of 30 to 50 words and idioms.

We explored seven strategies to select words. The goal of each was to identify words that appear to be used differently by different authors beyond what we would expect to observe by chance. To quantify the chances of observing variations or the function of variations of a given magnitude, we required a good model for word usage. This is where our Poisson and Negative-Binomial models once again came into play. For a word w, which was used X_{1w} and X_{2w} times by authors one and two, respectively, we compute

$$\Delta_w^2 = \frac{(X_{1w} - X_{2w})^2}{X_{1w} + X_{2w}}.$$

If the two authors used a word in a similar fashion we would expect to see small values of Δ^2. The larger the observed value, the stronger the indication we have that the occurrence patterns of a word across the texts is capturing elements of differential writing style.

https://doi.org/10.1017/S1049096507070874 Published online by Cambridge University Press

Analysis and Results

Some of our preliminary explorations led to dead ends. For example, in some of the early analyses it appeared possible to classify perfectly Reagan speeches using abbreviations, punctuation, and the American and Canadian spellings of words such as "theatre" or "theater." Further inspection of the original documents revealed that Reagan spelled such words both ways, and that he was inconsistent in spelling and punctuation. In addition, his spelling was often changed from an original manuscript during retyping or editing. The original drafts were typed, both in the offices of Deaver & Hannaford and in the offices of Harry O'Connor (where they were recorded), by many different individuals who had different views of spelling, capitalization, and the like. Thus, these differences were not indicators of authorship and we did not use these features in our subsequent analyses.

Eventually, our word selection strategies found words and idioms more indicative of the character behind the words. For example, we discovered that Reagan preferred colloquial expressions such as "over and over," whereas his collaborators preferred that he say "continuous" and "continuously." Reagan would attack his opponent, Jimmy Carter, in subtle ways. He would rarely mention the name "Carter." Rather, he would comment on the wrongdoings of "the government" and on the bad choices of "Capitol Hill people."

Let us discuss in more detail the use of the word "Carter" and other contextual words for predicting authorship. Following Frederick Mosteller and David Wallace (1964; 1984) in their analysis of the *Federalist Papers*, we looked for non-contextual words, with a few possible exceptions, as discriminators. The rationale behind this choice was that Reagan would discuss various issues and we did not want to bias our predictions by anticipating what his favorite topics were. It is conceivable, however, that presidential candidates may base their campaigns on specific sets of issues that they tend to emphasize in their speeches (Schonhardt-Bailey 2005). In making decisions about contextuality, our earlier idea of Reagan's style based on the text of the Reagan vs. Carter presidential debate, notes, comments, and books about Reagan, played a role. The analysis of the word "Carter" is an example. Our original analysis of Reagan's style suggested that, as mentioned above, he would seldom talk about his opponent as "Mr. Carter," but instead would take a more subtle line of attack. Thus, when the word "Carter" passed severe testing to make sure that its differential use by Reagan and Hannaford was too marked to be the outcome of pure chance and was likely to capture some element of Hannaford's writing style, we did not discard it as contextual.

Some have argued that Reagan's writing style might be better captured by some of the idioms he used. Thus we extended our analysis to the study of successive words to discover that, for example, idioms such as "if we," "in our," "I'd like to," or "in America" identify Reagan's writing style beyond reasonable doubts. Furthermore, Reagan cared about the impact of public policy on the finances of the average "worker" and his "wife" and would clearly explain how much each measure was "worth" to the people and what could be expected "as a result." He would often play a "joke" or tell a "story" to make his points. During his campaign, Reagan promised to make America the "shining city on a hill" that the entire world had admired in the good old days. In his radio addresses, he would communicate how things were "supposed" to be, and how, "when it comes" to resolving small everyday "issues," the American way of dealing with "problems" went a long way toward making things right without the need for "government" interventions. "And of course," "Mr. Carter" and "Capitol Hill people" were just messing it all up.

We then checked the assumptions underlying our model. In particular, we studied the pair-wise independence of words, the independence of subsequent occurrences of the same word across the text, and the goodness of fit of Poisson and Negative-Binomial distributions as models of word counts.

In general, the independence assumptions (pair-wise independence of words and independence of words from positions in the text of the speeches) were two major shortcuts that we used in our models. These assumptions are crucial in that they enabled us to cut down to a feasible number the probabilities to be estimated in several cases. Even though pair-wise independence statistically held in our data, and in our analysis we mostly focused on non-contextual, high-frequency words, we would expect these assumptions not to be true in general. In particular, a more desirable model would account for some functional form of dependence, for example, "attraction and repulsion," among words. Further, the extent to which independence of words from positions in the text holds is questionable for the positions at the end of sentences, and somewhat questionable for the positions at the beginning of sentences. Nonetheless, because of our reliance on out-of-sample estimation of the prediction error, the results these assumptions lead to are not overstatements or misrepresentations. Rather, the assumptions relating to independence at worst result in poorer accuracy than what we might achieve if we captured dependence appropriately. Out-of-sample accuracies above 90% in all cases, and predicted authors for the unknown speeches stable across 21 possible scenarios, support the simplicity of our models.

The goodness-of-fit study indicated that the Negative-Binomial model is appropriate for word counts and semantic features counts. As a consequence, we based our preferred word selection scheme, based on thresholding the statistic Δ^2, and the likelihood of the data upon this model (see Table 2).

We chose the constants β underlying the prior distributions with the aim of mitigating the variations in the use of words that would play a role in the attribution of authorship. We ran our experiments for 21 sets of constants entailing possible scenarios that we identified as reasonable with two small studies on 90 and 120 words and on speeches drafted by Reagan and his collaborators.

The remarkable descriptive power of the Negative-Binomial model fully translated into predictive power. The predictions we obtained with the fully Bayesian Negative-Binomial model were very stable, both in terms of accuracy and estimated out-of-sample across 21 sets of constants and in terms of predicted authorship for the 312 unknown speeches. We provided separate models for the speeches given in 1975 and for those given from 1976 to 1979 and obtained stable and accurate predictions on speeches given in different years about various topics.

That we were able to find serious statistical cues in the texts of the speeches and thus distinguish Reagan's "own hand" from those of his collaborators is striking. This is especially true if we consider that a professional editorial assistant typically "enters the head" of his boss in order to be helpful in drafting speeches and other documents.

Predictions for the Drafts of Uncertain Authorship

We produced Negative-Binomial and Poisson predictions to resolve the attributions for those speeches of uncertain authorship. Furthermore, we produced predictions for the best off-the-shelf data-mining classifiers: the logistic regression and the naïve-Bayes classifier based on

https://doi.org/10.1017/S1049096507070874 Published online by Cambridge University Press

Table 2
Goodness of Fit of Poisson and Negative-Binomial Models for Pools of Words Selected with Different Methods[6]

	Poisson Model		Negative-Binomial Model		
Pool of Words	Hannaford (38 texts)	Reagan (75 texts)	Hannaford (38 texts)	Reagan (679 texts)	Reagan (75 texts)
50 highest frequency words	12 (50)	3 (50)	31 (50)	0 (50)	49 (50)
54 words in pool no. 1	4 (15)	0 (17)	14 (15)	2 (17)	13 (17)
21 features in pool no. 3	3 (21)	1 (21)	21 (21)	0 (21)	20 (21)
49 *n*-grams in pool no. 4	1 (12)	0 (14)	12 (12)	2 (14)	14 (14)
27 words in pool no. 6	1 (11)	0 (11)	10 (11)	1 (11)	11 (11)
31 words in pool no. 7	1 (5)	0 (3)	5 (5)	0 (3)	1 (3)
27 words in pool no. 9	0 (7)	0 (8)	7 (7)	2 (8)	8 (8)

the multinomial distribution. The parameter values were estimated at the posterior means using the information in the "known" texts for all sets of underlying constants, for the words selected with the Delta-Squared statistic, and for the semantic features. The magnitude of the log-odds of authorship entailed clear-cut predictions for the authorship of many of the unknown speeches. In addition, the bold agreement of several accurate classification methods, based on both the analysis of words and a semantic decomposition of the speeches, reinforced our confidence. As a measure of the goodness of our predictions, we present a three-way table that displays the degree of agreement among our classifiers on the speeches of uncertain authorship. We compose the predictions obtained from the fully Bayesian models using several sets of underlying constants with the predictions of the multinomial naïve-Bayes and the logistic regression classifiers. Notice that more speeches are assigned to Reagan by the logistic regression classifier than by the Bayesian models. Nonetheless, the three classifiers all agree on 207 out of 312 speeches (66.3%). We fine-tuned our models to perform differential predictions regarding the author of the text of speeches delivered in different epochs, given that in 1975 it is likely that collaborators other than Peter Hannaford may have contributed to the drafts (see Tables 3 and 4).

Table 3
Comparison of 311 Predictions under Three Models

	Poisson Full Bayes (β no. 1,4,8)			
	Hannaford		Reagan	
	Logistic Regression		Logistic Regression	
Multinomial Naïve Bayes	Hannaford	Reagan	Hannaford	Reagan
Hannaford	53	31	26	8
Reagan	8	10	21	154

Post-Analysis External Validation

Following completion of the analyses reported above, Annelise Anderson found more yellow notepads written in Reagan's hand including drafts of three more radio speeches, namely "FBI," "London 1," and "London 2." The Poisson and Negative-Binomial classifiers had correctly predicted that Reagan was the most likely author of those drafts, independently of reasonable variations in the underlying constants and prior odds of authorship (see Table 5).

More recently, we re-examined the unknown speeches that were likely to be drafted by Reagan according to most, and in some cases all, of our classifiers. In 10 of them he comments on the Strategic Arms Limitation Talks (SALT) or the degree of deterrence that would be possibly induced by strategic nuclear bombs or intercontinental ballistic missiles. The accuracy of our methodology, on the basis of out-of-sample experiments as well as on the examples confirmed via the newly discovered yellow notepads, adds support to our confidence in the predictions for these papers. The speeches about SALT and deterrence we so identified would thus be evidence of Reagan's pre-presidential commitment to nuclear arms reductions, a commitment he would act upon as president.

Relevance of Our Findings

Familiarity with Reagan's pre-presidential writings has led to a more complete understanding of his policies as president and will most likely continue

Table 4
Prediction Accuracy of Four Models, Estimated Using 1,000 Out-of-Sample Experiments

	Logistic Regression	Naïve-Bayes Multinomial	Full-Bayes Poisson	Full Bayes Neg.-Bin.
1975	89.0% (7.6%)	91.5% (3.6%)	88.7% (5.5%)	93.3% (5.8%)
1976–79			88.0% (3.5%)	95.3% (1.1%)

https://doi.org/10.1017/S1049096507070874 Published online by Cambridge University Press

Table 5
Posterior Log Odds of Reagan's Authorship for the Three Handwritten Notepads Recently Discovered by Annelise Anderson[7]

Speech	$\beta(1)$	$\beta(2)$	$\beta(3)$	$\beta(6)$	$\beta(7)$	$\beta(12)$	$\beta(16)$	$\beta(20)$
F.B.I.								
Words (Poisson)	16.99	17.65	16.45	17.70	16.03	17.06	17.14	17.70
Words (N.B.)	15.05	15.63	14.57	15.66	14.22	15.05	15.21	15.74
Semantics (N.B.)	0.13	0.08	0.07	0.06	0.10	0.13	0.11	0.13
London I								
Words (Poisson)	42.02	43.51	40.55	43.72	39.61	42.03	42.50	43.80
Words (N.B.)	21.37	22.13	20.59	22.20	20.11	21.32	21.61	22.23
Semantics (N.B.)	0.48	0.40	0.41	0.48	0.37	0.45	0.46	0.41
London II								
Words (Poisson)	14.91	15.41	14.39	15.48	14.04	14.93	15.02	15.54
Words (N.B.)	5.52	5.73	5.31	5.75	5.22	5.53	5.56	5.74
Semantics (N.B.)	0.36	0.37	0.37	0.33	0.35	0.31	0.39	0.36

to do so as scholars mine the ever-increasing body of pre-presidential materials being released to the public at the Ronald Reagan Presidential Library. Attention to Reagan's early writings may also illuminate the important issue of how his anti-classical position on strategies of deterrence contributed to the peaceful (non-nuclear) ending of the Cold War. Our word-count study suggests that President Reagan may not have had a scriptwriter at all; he was much more the author of his words than many have believed.

Notes

*The authors would like to thank Jared Cohon, president of Carnegie Mellon University, for suggesting this study, Robyn Dawes, Charles J. Queenan, Jr., University Professor in Social and Decision Sciences at Carnegie Mellon, and Annelise Anderson, research fellow at the Hoover Institution, for advice and analysis throughout this project.

1. Skinner and the Andersons co-edited a second book of Reagan's handwritten radio commentaries, *Reagan's Path to Victory: The Shaping of Ronald Reagan's Vision: Selected Writings* (Northampton, MA: Free Press, 2004). In 2003, Skinner and the Andersons also co-edited a collection of approximately 1,000 of Reagan's letters based on more than 70 years of his life, *Reagan, A Life in Letters* (Northampton, MA: Free Press).

2. For instance, in a *Time* cover story on *Reagan, A Life in Letters*, Michael Duffy and Nancy Gibbs (2003) wrote: "Future scholars may argue with the substance of Reagan's principles but not with their pedigree, for now they will have a paper trail of the kind historians can only dream. . . ." See also Brooks (2001), Cannon (2001), and Safire (2000).

3. For a thorough discussion of this issue we refer to Airoldi et al. (2006).

4. For details of this procedure we refer to Airoldi et al. (2006).

5. Multiple overlapping boxes are suggestive of the 21 sets of underlying constants, β, that were used in the fully Bayesian strategy to perform sensitivity analysis.

6. We quote the number of words that fit the Poisson and Negative-Binomial profiles in each collection of texts, using the corresponding p-values obtained with two-sample Kolmogorov-Smirnov tests. Rare words were discarded; in brackets we quote the actual number of words in each pool for which a p-value was computed. The columns labeled "Reagan (75 texts)" contain the average results of tests over 100 samples of 75 texts each. (This is to avoid statistical significance due to large sample size, but not substantive.)

7. The different columns correspond to different sets of underlying constants for the Poisson and Negative-Binomial models of words and semantic features described in the main text.

References

Airoldi, E. M., A. G. Anderson, S. E. Fienberg, and K. K. Skinner. 2006. "Who Wrote Ronald Reagan's Radio Addresses?" *Bayesian Analysis* 1: 289–320.

Brooks, D. 2001. "Reagan Was a Reaganite," *New York Times Book Review*, January 28, 5.

Cannon, L. 2001. "A Way with Words," *Washington Post Book World*, February 4, 1–3.

Carlin, B. P., and T. A. Louis. 2000. *Bayes and Empirical Bayes Methods for Data Analysis, 2nd ed*. Danvers, MA: CRC Press.

De Morgan, A. 1872. *Budget of Paradoxes*. London: Longmans, Green.

Duffy, M., and N. Gibbs. 2003. "The Real Reagan: Think you know what made him tick? His letters may surprise you," *Time*, September 29, 55.

FitzGerald, F. 2000. *Way Out There in the Blue: Reagan, Star Wars, and the End of the Cold War*. New York: Simon and Schuster.

George, A. L., and J. L. George. 1998. *Presidential Personality and Performance*. Oakforest, IL: Westview Press.

Greenstein, F. I. 1990. "Ronald Reagan: Another Hidden-Hand Ike?" *PS: Political Science and Politics* 23 (1): 10.

Laver, M., K. Benoit, and J. Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97: 311–31.

Lettow, P. V. 2005. *Ronald Reagan and His Quest to Abolish Nuclear Weapons*. New York: Random House.

Mendenhall, T. C. 1887. "The Characteristic Curves of Composition." *Science* 11: 237–49.

Mitchell, T. 1997. *Machine Learning*. New York: McGraw-Hill.

Mosteller, F., and D. L. Wallace. 1964. *Inference and Disputed Authorship: The Federalist*. Boston: Addison-Wesley.

———. 1984. *Applied Bayesian and Classical Inference: The Case of The Federalist Papers*. New York: Springer-Verlag.

Safire, W. 2000. "Reagan Writes," *New York Times Magazine*, December 31, 38–42.

Schonhardt-Bailey, C. 2005. "Measuring Ideas More Effectively: An Analysis of Bush and Kerry's National Security Speeches." *PS: Political Science and Politics* 38 (4): 701–11.

Skinner, K. K., A. Anderson, and M. Anderson, eds. 2001. *Reagan, in His Own Hand: The Writings of Ronald Reagan that Reveal his Revolutionary Vision for America*. Northampton, MA: Free Press.

———. 2003. *Reagan, A Life in Letters*. Northampton, MA: Free Press.

———. 2004. *Reagan's Path to Victory: The Shaping of Ronald Reagan's Vision: Selected Writings*. Northampton, MA: Free Press.

https://doi.org/10.1017/S1049096507070874 Published online by Cambridge University Press