

The Non-Traditional Case for the Authorship of the Twelve Disputed "Federalist" Papers: A Monument Built on Sand?

Joseph Rudman (jr20@andrew.cmu.edu)
Carnegie Mellon

Introduction

This paper discusses the controversy over the authorship of twelve of the "Federalist" papers as seen and studied by over twenty non-traditional authorship attribution practitioners. The "Federalist" papers were written during the years 1787 and 1788 by Alexander Hamilton, John Jay, and James Madison. These 85 propaganda tracts were intended to help get the U.S. Constitution ratified. They were all published anonymously under the pseudonym, "Publius." The general consensus of traditional attribution scholars (although varying from time to time) is that Hamilton wrote 51 of the papers, Madison wrote 14, Jay wrote 5, while 3 papers were written jointly by Hamilton and Madison, and 12 papers have disputed authorship — either Hamilton or Madison.

In 1964, Frederick Mosteller and David Wallace, building on the earlier unpublished work of Frederick Williams and Frederick Mosteller, published their non-traditional authorship attribution study, "Inference and Disputed Authorship: The Federalist." It is arguably the most famous and well respected example from all of the non-traditional attribution studies. It is the most statistically sophisticated non-traditional study ever carried out. There even has been a 40 page paper explicating the statistical techniques of the Mosteller and Wallace study (Francis). Since then, hundreds of papers have cited the Mosteller and Wallace work and over two dozen non-traditional attribution practitioners have analyzed and/or conducted variations of the original study.

These practitioners wanted to test their statistical approaches against the Mosteller and Wallace touchstone study. Mosteller and Wallace set the boundary conditions for the subsequent work — e.g., not using the Jay articles as a control. Their experimental design and overall report is never questioned. Most of these later practitioners do not select or prepare the input text as rigorously as Mosteller and Wallace — whose own selection

and preparation was not as rigorous and complete as it should have been.

Text Selection

(1) "Federalist" Papers

This section discusses the way the Federalist papers were originally published (76 in newspapers and 8 in the book compilation) and which editions the practitioners chose for their non-traditional studies — how 84 papers became 85 and how some papers had different numbers in different editions. The effect that the lack of Hamilton and Madison holographs had on the studies is discussed. The choice of edition has the potential of profoundly changing the results of the studies.

Project Gutenberg Etexts are usually created from multiple editions, all of which are in the Public Domain in the United States, unless a copyright notice is included. Therefore we do NOT keep these books in compliance with any particular paper edition, usually otherwise.

(Front Material of Gutenberg Etext #1404)

The compounding problem of down-loading texts via the internet is explicated — e.g., one of the texts includes every variant of every paragraph. It is shown why none of the Federalist studies used a 'valid' text of the Federalist papers. The question, "Does this incorrect input data invalidate the final 'answer?'" is discussed.

(2) The Control Texts

(a) The "Known" Hamilton Sample

This sample cannot contain questionable Hamilton writings. This sample must also fulfill the other criteria of a valid sample — e.g., same genre, same constricted time frame. There also should be a sub-set of this sample set aside for later analysis in order to guard against the charge of cherry picking the style-markers. This is not the same as the Mosteller and Wallace "training sample."

(b) The "Known" Madison Sample

In addition to discussing the way the Madison sample was constructed, what was said about the Hamilton sample will be applied here.

Does the lopsided number of Hamilton papers over Madison papers (51 to 14) pose a problem for the studies? Were the Hamilton and Madison control texts from outside the Federalist papers chosen correctly? Why are these "outside" controls not used by most of the other practitioners? This section goes on to

discuss the control problems that arose with the Mosteller and Wallace study and have been perpetuated through the subsequent studies. This section also discusses the other control problems introduced in these studies.

Text Unediting, De-editing, and Editing

The cumulative effect of NEARLY A THOUSAND SMALL CHANGES [emphasis mine] has been to improve the clarity and readability of the text without changing its original argument.

(Scigliano, lii)

(1) The "Little Book of Decisions"

In the Mosteller and Wallace study, a "little book of decisions" is mentioned. This "book," originally constructed by Williams and Mosteller, contained an extensive list of items that Mosteller and Wallace unedited, de-edited, and edited before beginning the statistical analysis of the texts — items such as quotations and numerals. Unfortunately, neither Williams and Mosteller nor Mosteller and Wallace published the contents of this "little book of decisions" and only mention five of their many decisions in the published work. [Mosteller and Wallace 7, 16, 38-41] The little book has been lost and cannot be recovered or even reconstructed [Mosteller]. This paper goes on to discuss the many ramifications of the "little book" on their study and the subsequent studies. Also, how the loss of the "little book" casts a shadow of "scientific invalidity" over the Mosteller and Wallace work — i.e., it cannot be replicated. Their "little book" was not used by any of the following studies — making meaningful comparisons moot.

(2) Other Decisions

This section goes on to list many of the unediting, de-editing, and editing items that need to be considered. It lists several of the mistakes made by the many practitioners and what these mistakes mean to the validity of the studies (e.g.):

- (a) Wrong letters
- (b) Quotes — e.g., 131 words of Federalist 5 are a quote from Queen Ann, 334 words of Federalist 9 are a quote from Montesque
- (c) Footnotes — the author's and the editors'
- (d) Numbers
- (e) Foreign languages
- (f) Spelling
- (g) Homographic forms
- (h) Contracted forms

- (i) Hyphenation
- (j) Word determination
- (k) Disambiguation
- (l) Editorial intervention — internal (e.g., Hamilton on Madison) and external (e.g., from the first newspaper copy editor to present day editors)

Conclusion

(1) Acceptance of Results by Non-Traditional Practitioners

Are practitioners (statisticians and non-statisticians) so blinded by the statistical sophistication that the other elements of a valid non-traditional authorship study are ignored?

(2) Acceptance of Results by History Scholars

Do professional historians accept, deny, or show indifference to the body of work that supports the Mosteller and Wallace study? Why did I spend hours searching for a Mosteller and LAWRENCE study of the Federalist papers?

(3) Do the multiple flaws in all of these non-traditional studies invalidate the results.

Is the case put forth by Mosteller and Wallace and buttressed by the other non-traditional practitioners nothing but a "Monument" built on sand? What effect does showing the flaws in the Federalist studies have on non-traditional studies in general — i.e., if the best is suspect, what about the rest!

Bibliography

Adair, Douglass. "The Authorship of the Disputed Federalist Papers." *The William and Mary Quarterly* 1.2 Part I and 1.3 Part II (1944): 97-122 and 235-264.

Avalon Project. Yale Law School. 97-122 Ind 235-264. Accessed 13 February 2004, 10:30AM. <<http://www.yale.edu/lawweb/avalon/>>

Bosch, Robert A., and Jason A. Smith. "Separating Hyperplanes and the Authorship of the Disputed Federalist Papers." *The American Mathematical Monthly* 105.7 (1998): 601-607.

Bourne, E.G. "The Authorship of the Federalist." *The American Historical Review* 2.3 (1897): 443-460.

- Collins, Jeff, et al. "Detecting Collaborations in Text: Comparing the Authors' Rhetorical Language Choices in the Federalist Papers." *Computers and the Humanities* 38.1 (2004): 15-36.
- Cooke, Jacob E., ed. *The Federalist*. Cleveland: Meridian Books (The World Publishing Company), 1956.
- Davis, George. "RE: Gutenberg edition of Federalist." Private E-mail, 20 November 2003 18:46:51.
- Engeman, Thomas S., et al., ed. *The Federalist Concordance*. Middletown, Connecticut: Wesleyan University Press, 1980.
- Farrington, Jill. *Analysing for Authorship*. Cardiff: The University of Wales Press, 1966.
- Farrington, Michael G., and Andrew Q. Morton. "Fielding and the Federalist." *Department of Computing Science Research Report* (1990/R6).
- Forsyth, Richard S. "Stylistic Structures: A Computational Approach to Text Classification." Diss. University of Nottingham, 1995.
- Francis, Ivor S. "An Exposition of a Statistical Approach to the Federalist Dispute." *The Computer and Literary Style*. Ed. Jacob Leed. Kent Ohio: Kent State University Press, 1966. 38-78.
- Fung, Glenn. "The Disputed Federalist Papers: SVM Feature Selection via Concave Minimization." *Proceedings of the 2003 Conference on Diversity in Computing*. Atlanta, Georgia, 2003. 42-46.
- Fung, Glenn. *CS 635 Project*. Spring Semester 1999. Accessed 2004-11-09. <<http://www.cs.wisc.edu/~gfung/GSVMFP.ps>>
- Fung, Glenn, and Olvi L. Mangasarian. "The Disputed Federalist Papers: SVM Feature Selection via Concave Minimization." Paper delivered at the CSNA 2002 Conference, Madison, Wisconsin. 15 June 2002.
- Hamilton, Alexander, et al. Ed. Robert Scigliano. *The Federalist: A Commentary on the Constitution of the United States*. New York: The Modern Library (Random House), 2000.
- Hart, Michael. "RE: Gutenberg edition of Federalist." Private E-mail, 21 November 2003 12:59:08.
- Hilton, Michael L., and David I. Holmes. "An Assessment of Cumulative Sum Charts for Authorship Attribution." *Literary and Linguistic Computing* 8.2 (1993): 73-80.
- Holmes, David I., and Richard S. Forsyth. "The Federalist Revisited: New Directions in Authorship Attribution." *Literary and Linguistic Computing* 10.2 (1995): 111-127.
- Khmelev, Dimitri V., and Fiona J. Tweedie. "Using Markov Chains for Identification of Writers." *Literary and Linguistic Computing* 16.3 (2001): 299-307.
- Kjell, Bradley. "Authorship Determination Using Letter Pair Frequency Features with Neural Network Classifiers." *Literary and Linguistic Computing* 9.2 (1994): 119-124.
- Kjell, Bradley, et al. "Discrimination of Authorship Using Visualization." *Information Processing & Management* 30.1 (1994): 141-150.
- Martindale, Colin, and Dean McKenzie. "On the Utility of Content Analysis in Author Attribution: The Federalist." *Computers and the Humanities* 29 (1995): 259-270.
- McColly, William, and Dennis Weier. "Literary Attribution and Likelihood-Ratio Tests: The Case of the Middle English Pearle-Poems." *Computers and the Humanities* 17 (1983): 65-75.
- Merriam, Thomas. "An Experiment with the Federalist Papers." *Computers and the Humanities* 23.3 (1989): 251-254.
- Mitchell, Ann F.S., and Clive D. Payne. "A Conservative Confidence Interval for a Likelihood Ratio." *Journal of the American Statistical Association* 66.336 (1971): 861-866.
- Mosteller, Frederick, and David L. Wallace. "Notes on an Authorship Problem." *Proceedings of a Harvard Symposium on Digital Computers and their Applications*. Cambridge, Massachusetts: Harvard University Press, 1962. 163-197.
- Mosteller, Frederick, and David L. Wallace. "Inference in an Authorship Problem. A Comparative Study of Discrimination Methods Applied to the Federalist Papers." *Journal of the American Statistical Association* 58 (1963): 275-309.
- Mosteller, Frederick, and David L. Wallace. *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. New York: Springer-Verlag, 1984.
- Pennebaker, James W. "The Federalist." Unpublished preliminary work.
- Pennebaker, James W. "[no title]." Private E-mail, Wednesday 09 July 2003, 14:45:34.
- Pennebaker, James W. "[no title]." Private E-mail, Wednesday 09 July 2003, 15:32:59.
- Piaia, Jesse. "[For Frederick Mosteller]." Private E-mail, Tuesday 22 July 2003, 10:57:38.
- Piaia, Jesse. "[For Frederick Mosteller]." Private E-mail, Tuesday 22 July 2003, 11:48:04.
- Project Gutenberg*. Accessed 2003-09-30. <<http://promo.net/pg/>>
- Rokeach, Milton, et al. "A Value Analysis of the Disputed Federalist Papers." *Journal of Personality and Social Psychology* 16.2 (1970): 245-250.
- Roland, Jon. "RE: The Federalist on constitution.org." Private E-mail, 11 September 2003, 10:24:36.

- Rudman, Joseph. "Unediting, De-Editing, and Editing in Nontraditional Authorship Attribution Studies: With an Emphasis on the Canon of Daniel Defoe." *Papers of the Bibliographical Society of America* 99:1 (March 2005).
- Sarndal, Carl-Erik. "On Deciding Cases of Disputed Authorship." *Applied Statistics* 16.3 (1967): 251-268.
- Stamatatos, E., N. Fakotakis, and G. Kokkinakis. "Text Genre Detexction Using Common Word Frequencies." *COLING 2000: Proceedings of the 18th International Conference on Computational Linguistics*. 2000. 808-814.
- Stamatatos, E., N. Fakotakis, and G. Kokkinakis. "Computer-Based Authorship Attribution Without Lexical Measures." *Computers and the Humanities* 35 (2001): 193-214.
- Tankard, Jim. "The Literary Detective." *BYTE* 11.2 (1986): 231-238.
- Tweedie, Fiona J., S. Singh, and D.I. Holmes. "Neural Network Applications in Stylometry: The Federalist Papers." *Computers and the Humanities* 30.1 (1996): 1-10.
- Wachal, Robert Stanley. *Linguistic Evidence, Statistical Inference, and Disputed Authorship*. Dissertation, University of Wisconsin, 1966.
- Waugh, Sam, Anthony Adams, and Fiona Tweedie. "Computational Stylistics Using Artificial Neural Networks." *Literary and Linguistic Computing* 15.2 (2000): 187-197.
- Yang, Albert C.C., et al. "Information Categorization Approach to Literary Authorship Disputes." *PHYSICA A* ().
constitution.org. Accessed 9-30-03. <<http://constitution.org/fed/feder00.htm>>