



2012

Stylometric Analyses of the Book of Mormon: A Short History

Matthew Roper

Paul J. Fields

G. Bruce Schaalje

Follow this and additional works at: <https://scholarsarchive.byu.edu/jbms>

BYU ScholarsArchive Citation

Roper, Matthew; Fields, Paul J.; and Schaalje, G. Bruce (2012) "Stylometric Analyses of the Book of Mormon: A Short History," *Journal of Book of Mormon Studies*: Vol. 21 : No. 1 , Article 4.

Available at: <https://scholarsarchive.byu.edu/jbms/vol21/iss1/4>

This Feature Article is brought to you for free and open access by the Journals at BYU ScholarsArchive. It has been accepted for inclusion in *Journal of Book of Mormon Studies* by an authorized editor of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.



NEAL A. MAXWELL INSTITUTE
FOR RELIGIOUS SCHOLARSHIP

BRIGHAM YOUNG UNIVERSITY • PROVO, UTAH

Title Stylometric Analyses of the Book of Mormon: A Short History

Author(s) Matthew Roper, Paul J. Fields, and G. Bruce Schaalje

Reference *Journal of the Book of Mormon and Other Restoration Scripture* 21/1 (2012): 28–45.

ISSN 1948-7487 (print), 2167-7565 (online)

Abstract The abundance of skeptical theories about who wrote the Book of Mormon has led many scholars to seek scientific data to discover the answer. One technique is stylometry. Having first been developed in the 1850s, stylometry seeks to find the “wordprint” of a text. Although these stylistic studies are not as accurate as a human’s fingerprint, they can give researchers a good idea either of differences in style between authors or of who might have written a text from a list of possible authors. Beginning in the 1960s individuals have completed four major stylometric studies on the Book of Mormon, studies that varied in both findings and quality of research. In addition to these four studies, this article presents a fifth study—using extended nearest shrunken centroid (ENSC) classification—that incorporates and improves on the earlier research.

STYLOMETRIC ANALYSES OF THE BOOK OF MORMON: A SHORT HISTORY



Photograph by Mark Philbrick.

FROM THE EDITOR:

What value do computerized studies of author styles contribute to the polemics and irenics that seem to perpetually swirl around the Book of Mormon? In this article, authors Roper, Fields, and Schaalje take a few short steps back to take a long look at what such studies can and cannot contribute, including the latest twist, nearest shrunken centroid (NSC) classification. The authors present eight serious flaws with the NSC study and then offer the results of their recent study using extended nearest shrunken centroid (ENSC) classification, which overcomes those flaws. Long-time readers of FARMS publications and those of the Neal A. Maxwell Institute will enjoy this short history.

Claims about the authorship of the Book of Mormon have a history as long as the book has been around. To discredit Joseph Smith's description of the book's origin, skeptics started proposing theories about who had written it even before it was published in 1830.¹ In 1834 Eber Howe proposed the Spalding-Rigdon theory of Book of Mormon authorship,² which asserts that Sidney Rigdon plagiarized an unpublished fictional work by Solomon Spalding to produce the Book of Mormon. He made this assertion even though the Book of Mormon was printed before Rigdon joined the church. Similar allegations and variations on that theme continue today, despite solid historical evidence that the theory is a baseless fabrication.³ Another way to look for evidence that supports or does not support specific claims of authorship is to examine the writing styles in a text, specifically by identifying word-use patterns. In this article, we look at the strengths and weaknesses of various word studies that have attempted to determine who wrote the Book of Mormon. We conclude with the results of our own study of Book of Mormon authorship.

Stylometry

When reading a written text, a reader may often identify words and phrases that seem to ring with a familiar voice, such that he or she may say, "This sounds like it was written by Mark Twain (or Ernest Hemingway or William Shakespeare)." But this is a very subjective judgment. On the other hand, stylometry, also known as computational stylistics, is a method of authorship attribution that uses far less subjective criteria—namely, statistical techniques—to infer the authorship of texts based on writing patterns. It tries to describe an author's conscious and unconscious creative actions with quantifiable measures such as the frequency with which an author uses certain words or groupings of words.

Stylometric analysis is based on the fundamental premise that authors write with distinctive, repeated patterns of word use. According to English professor John Burrows, written texts have a particular style and inherently display the intellectual propensities of their authors.⁴ By identifying the word-use patterns

in a text of unknown or questioned authorship and then comparing and contrasting those patterns to the patterns in texts of known authorship, the similarities and dissimilarities between the textual patterns can provide supporting evidence for or contradicting evidence against an assertion of authorship.

Anonymous writing, plagiarism, and the consequent debates about the authorship of texts have a long history. . . . Ancient catalogs of Aristotelian writings disagree . . . as to which works Aristotle actually wrote.

Anonymous writing, plagiarism, and the consequent debates about the authorship of texts have a long history, perhaps extending back to the advent of writing itself. For example, three ancient catalogs of Aristotelian writings disagree with each other as to which works Aristotle actually wrote.⁵ The authorship of Shakespeare's plays has been a topic of extensive debate and research,⁶ as has the authorship of the biblical epistles historically attributed to the apostle Paul.⁷ In the sixteenth century in England and Wales, a series of anonymous religious writings known as the Martin Marprelate tracts generated a great deal of controversy, including speculation about their authorship.⁸ *Common Sense*, published anonymously by Thomas Paine in January 1776, was the most influential tract of the American Revolution and became an instant best seller, both in the colonies and in Europe. To promote ratification of the United States Constitution, eighty-five short essays signed with the pseudonym "Publius" were published during 1787–88 in various New York City newspapers. They were later reprinted collectively as *The Federalist*. Although it was revealed in 1807 that the essays had been written by Alexander Hamilton, James Madison, and John Jay, the specific authorship of twelve essays remained in dispute for over 150 years until statistical analyses would show strong support for Madison as their author.⁹

A Brief History of Stylometry

The use of statistical tools to test questions of authorship in such situations goes back at least to 1851, when mathematician Augustus de Morgan proposed using average word length to numerically characterize authorship style.¹⁰ In 1887 Thomas Mendenhall, a physicist, proposed that an author has a “characteristic curve of composition” determined by how frequently an author uses words of different lengths. He applied this approach to compare the works of Shakespeare and Francis Bacon, for example.¹¹ In 1888 William Benjamin Smith, a mathematician writing under the pseudonym Conrad Mascol, published two papers describing a “curve of style” based on average sentence lengths to distinguish authorial styles, which technique he applied to the Pauline

A few advances in stylometry were made in the first half of the twentieth century, but the most significant step was the landmark publication in 1964 of statisticians Frederick Mosteller and David Wallace.

Epistles.¹² Then in 1893, Lucius Sherman, a professor of English, found that average sentence length could be used as an indicator of changes in writing styles over time.¹³

A few advances in stylometry were made in the first half of the twentieth century, but the most significant step was the landmark publication in 1964 of statisticians Frederick Mosteller and David Wallace. In their study they innovatively applied Bayesian statistical principles to investigate the authorship of the twelve disputed essays in *The Federalist*.¹⁴ From the late 1980s to the early 2000s, John Burrows made seminal contributions to stylometric methodology. He introduced the “delta score” to measure word frequency differences among texts that varied by author, in genre, or even across time periods.¹⁵ His method is now considered a benchmark for authorship attribution studies. Burrows also started a trend of using principal components analysis in stylometry.¹⁶

Today, the field of stylometrics is growing rapidly due to the confluence of exponentially increasing computing power, ubiquitous availability of the

Internet, development of ultrahigh dimensional statistical tools, and advances in Bayesian statistics.

Limitations of Stylometry

Stylometry is a useful tool in authorship attribution, but several limitations are important to keep in mind when interpreting the results of a stylometric analysis. Although stylometry is sometimes referred to as wordprint analysis (implying that it is a linguistic equivalent to fingerprint analysis), it does not have the same identifying capability. The description of stylometry as verbal DNA is an even less applicable overstatement.¹⁷ With stylometrics there is no way to perform population studies to determine the general prevalence of word-use patterns. Consequently, all probability assessments in stylometrics are relative only to the specific authors and the texts included in the study.

Although a person’s fingerprint and DNA are unchangeably unique to that person, a writer is at liberty to adapt his or her style to a particular topic, audience, and genre; to use artistic license to try new styles or even imitate others’ styles; and to modify his or her own style over time as writing skills increase or falter. Shakespeare, for example, was famously diverse in his writing style—an ability that is one of the hallmarks of a great author and also one of the things that makes stylometry a challenging methodology to apply successfully.

Further, writing style is not singularly specific to a person. Stylometry can assess the similarity of writing styles among authors, but it cannot prove personal identification of an author. Not only is there variation in an author’s word-use patterns, but authors can write sufficiently unlike themselves and sufficiently like each other at times that there are not clear boundaries between them, leaving fuzzy areas where their styles can overlap. So even though an author’s style may be distinctive, it is not distinct enough to be considered unique to that author to the exclusion of all other authors in the world.

Stylometric characteristics can provide a general comparative description of an author’s style, but the writing style exhibited in a text is an indirect and uncertain measure of an author’s identity. Authorial style is indistinct enough that one can say only, “Based on these style characteristics, this text could have been written by author X, and it was more likely written by author X than by author Y.” Thus, stylom-

etry can assess the probability of similar *writing styles* among texts, but that is not the same as the probability of *authorship* of those texts. Stylometry is only one source of evidence to support a claim of possible authorship. Other evidence—such as historical and biographical evidence—becomes essential.¹⁸

In the context of what stylometry is and what it is not, let us now consider the applications of the stylometric analyses that have been made regarding the question of authorship of the Book of Mormon.

Stylometric Analyses of the Book of Mormon

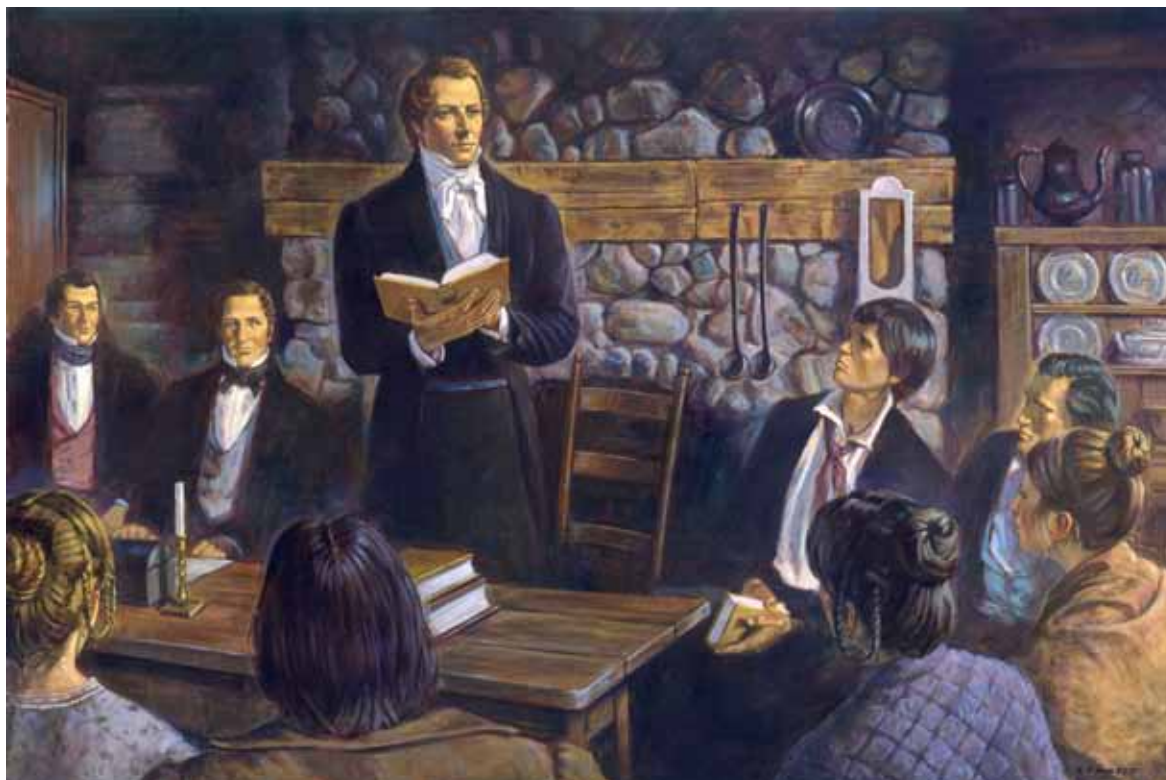
Since 1980, four major stylometric analyses of the Book of Mormon have been published—two by researchers at Brigham Young University,¹⁹ another by a doctoral student at Bristol Polytechnic,²⁰ and yet another by researchers at Stanford University.²¹ Each of these studies applied stylometry in different ways, seeking to address differing research questions, but all aimed at testing claims of Book of Mormon authorship.

The Larsen Study

Inspired by the Mosteller and Wallace study, three statisticians at Brigham Young University—

Wayne Larsen, Alvin Rencher, and Tim Layton—examined the frequencies of noncontextual words in a precedent-setting analysis of the Book of Mormon in 1980. Noncontextual words are function words that have a grammatical role forming the structure of a message, but they do not provide information about the message. These are words such as *a, an, but, however, the, to, with, without*, and so on. Mosteller and Wallace had shown that the way an author uses noncontextual words could be a means of characterizing the author's literary style independent of the author's message. For example, they found that Hamilton frequently used *enough* while Madison never used *enough* in his essays. Conversely, Madison frequently used *whilst*, and Hamilton never used that term. Mosteller and Wallace referred to such disparately used words as “markers” that could be used to distinguish between the writings of Hamilton and Madison—a process of authorial discrimination.

In the Larsen study the researchers carefully constructed 2,000-word text blocks for each of the major purported authors in the Book of Mormon. Then they tested whether the text blocks displayed evidence of a consistent style across the blocks, indicative of one author for all the texts, or whether



Organization of the church on 6 April 1830. Joseph Smith holds a copy of the Book of Mormon. © Robert T. Barrett.

there was evidence of differing styles, congruent with the claim that the Book of Mormon texts came from different writers.

Applying linear discriminant analysis²² based on the frequencies of noncontextual words occurring in each text block, the researchers used this technique to compare the authors specified internally in the Book of Mormon to a set of nineteenth-century authors external to the Book of Mormon. The statistical evidence of differences among the writings of the purported authors was overwhelming:

1. “Distinct authorship styles can be readily distinguished within the Book of Mormon.”
2. “The nineteenth-century authors do not resemble Book of Mormon authors in style.”²³

A summary plot of their findings in figure 1 shows how the texts form clusters for each of the four major authors identified internally in the Book of Mormon with a separate cluster for Joseph Smith as an external author; his personal writings were used in the comparison.

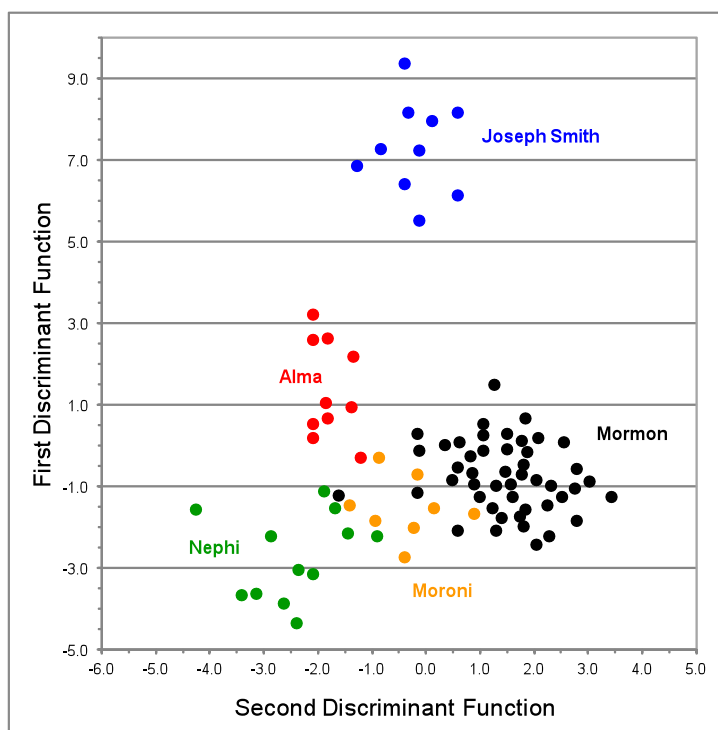


Figure 1. Text clusters of major Book of Mormon authors and Joseph Smith. Linear discriminant analysis indicates that the writing styles of the major Book of Mormon authors are distinguishable from each other and highly distinctive from Joseph Smith’s writing style.

We can see that the text blocks attributed to Nephi, Alma, Mormon, and Moroni in the Book of Mormon are consistently similar within authors (tight grouping of texts by author) but consistently different among purported Book of Mormon authors (distinct cluster for each author, with some overlap). Joseph Smith’s texts are clearly separated from the Book of Mormon texts.

There is, of course, no statistical way to prove that the actual authors for the specific text blocks were Nephi, Alma, Mormon, and Moroni. But whoever the authors were, each one consistently wrote within his or her same style, and the styles differed from each other. If one person wrote the whole Book of Mormon, he or she possessed an unusual and uncanny ability to write in different styles and to switch back and forth consistently between those styles.

Although somewhat overstated, it is hard to disagree with the Larsen study’s main conclusion that “our study has shown conclusively that there were many authors who wrote the Book of Mormon.”²⁴

The Hilton Study

Skeptical of, but intrigued by, the results of the Larsen study, John Hilton—a physicist at the Lawrence Livermore National Laboratory in California and later a researcher at Brigham Young University—decided in 1982 to test the reproducibility of the Larsen study results since a fundamental tenet of scientific research is that results of a study must be reproducible by other researchers. In so doing, Hilton took a different approach than Larsen.²⁵ Rather than using noncontextual word frequencies as stylistic features, Hilton used sixty-five noncontextual word-pattern ratios suggested by Andrew Morton,²⁶ a mathematician and religious studies scholar. Word-pattern ratios measure the rate of word use in four categories:

1. Specific words in key positions of sentences, e.g., “*the* as the first word of a sentence,”
2. Specific words adjacent to certain parts of speech, e.g., “*and* followed by an adjective,”
3. Collocations of words, e.g., “*and* followed by *the*,” and
4. Proportionate pairs of words, e.g., “*no* and *not*,” “*all* and *any*.”

		Number of Pairwise Rejections															
Text vs. Text		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Intra-Author Comparisons	Nephi vs. Nephi			1		1	1										
	Alma vs. Alma		1	1	1												
	Smith vs. Smith	1		2													
	Cowdery vs. Cowdery		1														
	Spalding vs. Spalding			1													
Inter-Author Comparisons	Nephi vs. Alma			1			2	2	1	1	1	1					
	Smith vs. Nephi					1				2		1	1	1			
	Smith vs. Alma				2	1	1		2								
	Cowdery vs. Nephi							1	1				2		1	1	
	Cowdery vs. Alma								4	1	1						
	Spalding vs. Nephi											1	1	1		1	2
	Spalding vs. Alma							3		2				1			
		Evidence of Similar Styles ←							→ Evidence of Different Styles								

Figure 2. Rejections of pairwise comparisons of texts from the Hilton study. Pairwise rejections fewer than seven of the possible sixty-five word pattern ratios in each text vs. text comparison indicate evidence of similar authorial style. The intra-author comparisons tend to show similar styles while the inter-author comparisons tend to show dissimilar styles.

Hilton’s idea was that these ratios might be minimally affected by unique phrases in the texts or by topic and genre differences among the texts and thus might be better detectors of an author’s unconscious word-use preferences. In agreement with Morton, Hilton reasoned that these word-pattern ratios would be useful since they provide a nonambiguous count, occur frequently, have common alternative expressions, and tend to be used habitually.

In addition, he developed a stylometric measure used to differentiate between any two texts based on the number of word-pattern ratios judged to be significantly different than expected (called rejections) between texts purportedly alleged to be written in the same authorial style. He calibrated and validated his method by applying it to texts of undisputed authorship from the 1800s and 1900s. He determined that seven or more rejections provided evidence of differences of writing style indicative of different authorship.

Using the oldest extant versions of the Book of Mormon—primarily the printer’s manuscript²⁷—he applied his procedure to 5,000-word blocks of text. This provided high reliability since in larger text blocks an author’s writing habits and stylistic propensities should assert themselves more strongly than in smaller texts. In compiling the text blocks,

he excluded quotations from the Bible and the distinctive phrase *and it came to pass*.

Hilton then made various comparisons among Book of Mormon texts attributed to Nephi and Alma and non-Book of Mormon texts known to have been authored by Joseph Smith, Oliver Cowdery, and Solomon Spalding. Specifically, he compared each author to the texts attributed to himself (intra-author comparisons) and then each author to every other author (inter-author comparisons). Figure 2 summarizes his results in tabular form. The first line of the table (Nephi vs. Nephi) indicates, for example, that there were three 5,000-word Nephi texts, and pairwise comparisons of these texts yielded two, four, and five rejections for tests of the sixty-five word-pattern ratios. Further, in comparing six sets of texts by Cowdery and Alma, four showed seven pairwise rejections, one showed eight, and the other nine, thus showing their dissimilarity. In figure 2, the intra-author comparisons show evidence of similar style, while the inter-author comparisons show evidence of dissimilar styles.

The most important result was that all of the Nephi, Alma, Smith, Cowdery, and Spalding texts are each consistent within themselves but distinctly different from one another. Thus, the evidence from the Hilton study argues strongly against the idea that

Joseph Smith, Oliver Cowdery, or Solomon Spaulding could be the author of the Nephi or Alma texts. Hilton concluded:

We show that it is statistically indefensible to propose Joseph Smith or Oliver Cowdery or Solomon Spaulding as the author of the 30,000 words from the Book of Mormon manuscript texts attributed to Nephi and Alma. Additionally these two Book of Mormon writers have wordprints unique to themselves and measure statistically independent from each other in the same fashion that other uncontested authors do. Therefore, the Book of Mormon measures [as being] multiauthored, with authorship consistent to its own internal claims.²⁸

Hilton's findings were congruent with the Larsen findings. In 2006 these results were reproduced again by researchers at Utah State University using generalized discriminant analysis—an extension of the linear discriminant analysis used in the Larsen study.²⁹

The Holmes Study

Not all Book of Mormon stylometric studies have reached the same conclusion as Larsen and Hilton.³⁰ For his doctoral dissertation at Bristol Polytechnic in 1985, David Holmes—now at the College of New Jersey but previously a professor at the University of the West of England—carried out a stylometric analysis of the Book of Mormon and related texts based on five measures of vocabulary richness.³¹ As stylistic features, Holmes computed a standardized measure of words used once in the text (R), a standardized measure of words used twice (V_2/V), a Poisson-based measure of lexical repetitiveness (K), and two estimated parameters of the Sichel distribution (α and θ)—a theoretical distribution to model word frequencies in writing. The first three measures were calculated for the total vocabulary in the texts, while the last two were calculated for nouns only.

His motivation was his impression at the time that vocabulary richness was a “particularly effective measure for discrimination between writers.”³² Holmes used the 1980 editions of the Book of Mormon, the Doctrine and Covenants, and the Book of Abraham from the Pearl of Great Price; the book of Isaiah from the King James Bible; and diaries and histories written or prepared by Joseph Smith between 1838 and 1843. Ignoring genre (doctrinal dis-

course versus historical narrative), Holmes extracted fourteen approximately 10,000-word blocks assigned to six Book of Mormon authors, divided sections 1 through 51 of the Doctrine and Covenants into three 10,000-word blocks, combined the writings of Joseph Smith into three 6,000-word blocks, included the Book of Abraham as one text, and extracted three 12,000-word blocks from Isaiah.

As illustrated in figure 3, Holmes found that the Joseph Smith texts clustered together, the Isaiah texts clustered together, and all but three of the other texts clustered together.

Holmes concluded from this that he had definitively shown that the writings of Mormon, Lehi, Nephi, Jacob, and Moroni were not stylometrically different. He stated, “There appears to be no real difference between Alma’s richness of vocabulary and Mormon’s richness of vocabulary, . . . a conclusion in direct contradiction to the findings of Larsen.” He continued, “This study has therefore not found any evidence of multiple authorship within the *Book of Mormon* itself,” to which he added, “We may consider the *Book of Abraham*, the purported authors of the *Book of Mormon* and Joseph Smith’s revelations to be of similar style, therefore, with all the implications that this may have for Mormon doctrine.”³³

The first part of Holmes’s statement is *prima facie* false since the Larsen study utilized noncontextual word frequencies and did not include any findings about vocabulary richness. The rest of the statement is an example of the classic fallacy *argumentum ad ignorantiam*: “I did not find a difference so there must not be a difference.” When a researcher does not find evidence of an effect, he or she can only say, “I did not find evidence of an effect.” The researcher cannot say, “Therefore, the effect does not exist.” The effect could still exist; the researcher simply did not find it. In addition, Holmes overgeneralized the usefulness of his methodology by failing to recognize that the successful application of a technique in one instance does not indicate that it is useful in all instances.³⁴ Even if a method found a large difference in one instance does not mean the method will find smaller differences in other cases. A method’s ability to find small differences that in fact exist is referred to by statisticians as the method’s power.

Subsequent research by Schaalje, Hilton, and Archer has shown that Holmes’s stylistic measures

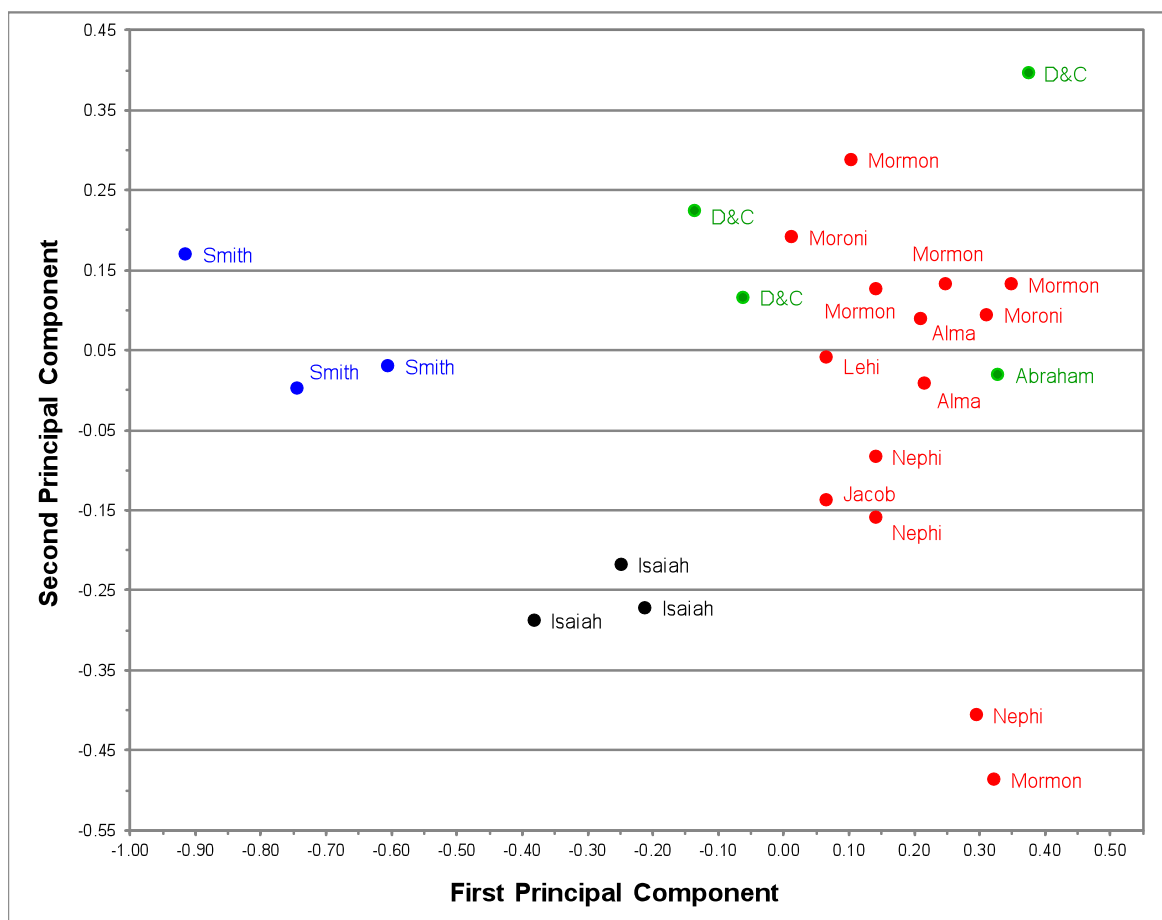


Figure 3. Principal components analysis plot based on Holmes’s vocabulary richness measures. Although texts from Joseph Smith and Isaiah are easily distinguishable from Book of Mormon, Doctrine and Covenants, and Pearl of Great Price texts, Holmes’s method could not distinguish among the purported authors within the Book of Mormon nor in comparison to the other scriptural texts.

have low power and are consequently weak discriminators of authorship.³⁵ For example, when testing texts of undisputed authorship by Samuel Clemens (Mark Twain) and Samuel Johnson (a British author and lexicographer), among others, correct classification rates were 96% using noncontextual word frequencies, 92% for noncontextual word-pattern ratios, but only 23% for vocabulary richness measures. Similar results were obtained consistently in other tests on sets of texts from novels (translated from German into English), the Book of Mormon texts (translated from an unknown ancient language into English), and the King James New Testament (translated from Greek into English). Later, in a reanalysis of *The Federalist* essays, Holmes himself found vocabulary richness measures to be comparatively less effective discriminators of authorship than noncontextual word frequencies.³⁶

The skepticism of Schaalje, Hilton, and Archer toward the effectiveness of Holmes’s vocabulary richness technique has been borne out in a more recent study by David L. Hoover:

Despite the attractiveness of measures of vocabulary richness, and despite the fact that they are sometimes effective in clustering texts by a single author and discriminating those texts from other texts by other authors, such measures cannot provide a consistent, reliable, or satisfactory means of identifying an author or describing a style. There is so much intratextual and intertextual variation among texts and authors that measures of vocabulary richness should be used with great caution, if at all, and should be treated only as preliminary indications of authorship, as rough suggestions about the style of a text or author, as characterizations of texts at the extremes of the range from richness to concentration. Perhaps their only significant usefulness is as an indicator of

what texts or sections of texts may repay further analysis by more robust methods. Unfortunately, the long-cherished goal of a measure of vocabulary richness that characterizes authors and their styles appears to be unattainable. The basic assumption that underlies it is false.³⁷

The results of the Holmes study certainly do not nullify the results of the Larsen and Hilton studies nor portend any grave implications for Mormon doctrine, as Holmes suggested. The Holmes study shows only that the Book of Mormon texts, although consistently distinct in terms of noncontextual word usage and word-pattern ratios, display similar vocabulary richness. This might reflect simply that the Book of Mormon texts are the work of a single translator, as Joseph Smith claimed, and thus were limited by his vocabulary.

The Jockers Study

The weakest of the four major Book of Mormon stylometric studies is presented in a recent paper by Matthew Jockers, Daniela Witten, and Craig Criddle³⁸—respectively an English lecturer, a statistics graduate student, and a civil engineering professor at Stanford University. Their study is innovative in that the statistical method they used was “nearest shrunken centroid classification” (NSC), a multivariate classification method based on Bayesian statistics developed for the classification of tumors in genomics research.³⁹

In statistics, shrinkage is a way to reduce the uncertainty about an estimated quantity by combining information from multiple sources in making the estimate. The more information that is included in making an estimate, the less uncertainty there will be about that estimate. A centroid is the center of a multidimensional cluster of data points. Think of it as the center of gravity of a disperse collection of related items with varying sizes. When applied to stylometry, the NSC method uses the stylistic characteristics (such as word frequencies) found in the texts of a set of candidate authors to create a rule for determining the authorship of unknown texts. That rule is then used to assign a text of questioned authorship to the author whose cluster of texts has the nearest centroid. The closer a test text of an unknown author is to the centroid of a known author’s texts, the greater the likelihood that the style of the test text matches the writing style of the known author.

Using Bayes’ theorem from statistics, the NSC

method updates initial probability estimates (called “prior probabilities”) to calculate final probability estimates (called “posterior probabilities”) based on newly obtained sample information. For example, without the sample information, the prior probability estimates would be that all candidate authors are equally likely to be the author of a text of unknown authorship. But after the writing style in the text (sample information) is compared with the writing style of each candidate author, the posterior probability estimates might show that one author is more likely the author of the text than the other candidates because of closer similarity of writing style. It is vitally important to note that NSC is a closed-set method, which means it assumes the set of candidate authors definitely includes the true author to the exclusion of any other possible candidates.

In the Jockers study, the researchers’ hypothesis was that the Book of Mormon is the collaborative work of multiple nineteenth-century authors. They specifically sought to find support for the Spalding-Rigdon theory. Therefore their set of candidate authors included text blocks by Solomon Spalding, Sidney Rigdon, Oliver Cowdery, and Parley P. Pratt. Biblical texts by Isaiah and Malachi (combined as one author) were included as a positive control, and contemporary nineteenth-century texts by Henry Longfellow and Joel Barlow were included as negative controls. The texts varied greatly in size, ranging from 114 to 17,797 words in length.

Even though chapter designations were not added to the Book of Mormon until 1879 (when all of their candidate authors were dead), Jockers chose to use the current chapter structure to define the test text blocks for the Book of Mormon, reasoning rather dubiously that the chapters might have been contributed individually by their panel of suspected authors and thus might provide evidence of “correct” authorship. The Book of Mormon chapters also varied widely in length from 95 to 3,752 words.

As stylistic features, Jockers used relative frequencies of the most common 110 words in the Book of Mormon that were used at least once by each purported author. From this list they removed four words that they felt were contextual in relating to biblical subject matter (*God, ye, thy, and behold*), but without justification they retained fifteen other contextual nouns: *children, day, earth, father, hand, king, land, man, men, name, people, power, son, time, and*

words. For some unknown reason they apparently wanted their definition of authorial style to include some lexical words—other than biblical-sounding words—rather than just function words.

The results of Jockers et al.’s application of NSC classification to assigning Book of Mormon chapters to their set of candidate authors are tabulated in figure 4.

There are eight serious flaws with the Jockers study methodology that render the results moot. First and most obviously, Joseph Smith was excluded as a candidate author, even though as the book’s translator he is the most likely author. His candidacy was considered in each of the previous studies. The Jockers researchers incorrectly claim that Joseph Smith could not be included because he frequently used scribes when preparing written documents and left inadequate samples of his personal writings. Dean Jessee has compiled a comprehensive set of Joseph Smith’s writings, many of which are holographic (written solely in his own hand).⁴⁰ Because NSC is designed to pick one of the members of a closed set of candidates, excluding Joseph Smith from the analysis seems like an attempt to stack the deck in favor of the Spalding-Rigdon theory authors.

Second, and even more important, the set of candidate authors for the Book of Mormon cannot reasonably be considered closed. To employ a

closed-set technique, a researcher must be assured by external evidence such as well-established, non-controversial historical information that all possible candidate authors have been identified and included. For *The Federalist* studies, there was no question that

Their candidate set cannot be considered to be comprehensive since the styles of the vast majority of Book of Mormon chapters differ markedly from the styles of any of Jockers et al.’s candidates.

the true author was included as a candidate. The question was only whether the writing style of a specific paper favored Hamilton or Madison; there were no other possible candidates. However, for the Book of Mormon the situation is not so simple—there is no substantiating historical or biographical information to justify a constrained set of candidates. In fact, the principal components plot of the Jockers study⁴¹ shown in figure 5 provides confirming evidence that their candidate set cannot be considered to be comprehensive since the styles of the vast majority of Book of Mormon chapters differ markedly from the styles of any of Jockers et al.’s candidates. Because of the dispersion of the data points (with very little overlap in the Book of Mormon clusters and the candidate authors’ clusters), it is obvious that the possibility of other authors than were allowed in the Jockers study must be included in an analysis of Book of Mormon writing styles.

Third, Jockers et al. assert that since twenty of the twenty-one chapters from Isaiah/Malachi were correctly attributed, “this is evidence for the effectiveness of NSC classification.”⁴² Yet they ignore the forty-two Book of Mormon chapters that are known *not to have been authored* by Isaiah and Malachi but that NSC mistakenly attributed to them. That means that NSC made *twice* as many incorrect attributions to Isaiah and Malachi as correct attributions. The statement Jockers et al. should have made is “This is evidence for the *ineffectiveness* of NSC classification.”

With two-thirds errors, this should have alerted Jockers et al. that their naïve application of NSC was producing unreliable results. This should have also

Nearest Shrunken Centroids (NSC) Classification Results	
Candidate Author	Percentage of Book of Mormon Chapters Assigned to an Author
Sidney Rigdon	39%
Isaiah/Malachi	26%
Solomon Spalding	22%
Oliver Cowdery	8%
Parley Pratt	4%
Henry Longfellow	1%
Joel Barlow	0%
Joseph Smith	Not Included

Figure 4. Percentage of Book of Mormon chapters assigned to each author by Jockers et al. based on nearest shrunken centroid (NSC) classification probability estimates, including Isaiah/Malachi as positive controls and Longfellow and Barlow as negative controls, but not including Joseph Smith.

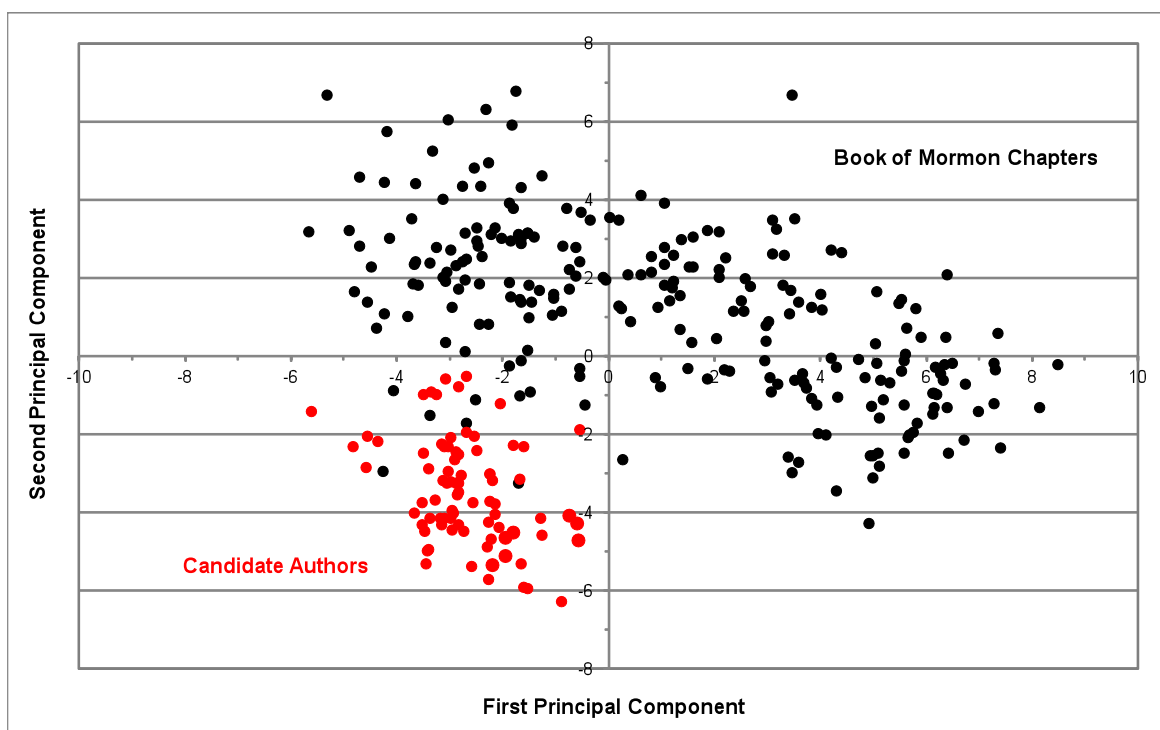


Figure 5. Principal components analysis plot for Jockers et al.’s data showing that the cluster for Book of Mormon chapters (black dots) is clearly separate from the cluster for candidate authors’ texts (red dots).

made them very suspicious that the ninety-three attributions to Rigdon must also be grossly overstated. If the same proportion of misattributions occurred for Rigdon as for Isaiah and Malachi, then the correct rate of attribution would be only about thirty-one chapters. As Jockers et al. point out, a mere random assignment of chapters would have resulted in thirty-four chapters attributed to each author.⁴³

Just as a stopped clock is right twice a day, NSC should be viewed as performing no better than attributing chapters by throwing darts at the list of candidate authors.

Jockers et al. should have realized therefore that the thirty-one chapters that *might have been* correctly attributed to Rigdon were only what would be expected by random assignment. Just as a stopped clock is right twice a day, NSC should be viewed as performing no better than attributing chapters by throwing darts at the list of candidate authors.

Fourth, even though the NSC method can identify the cluster of texts a test text is relatively *closest*

to, that does not mean it is *close* in an absolute sense. The test text and the closest cluster could still be a great distance apart. This would allow for the possibility that an excluded author is actually closer. As an analogy, let us ask the question, “Considering the cities New York, Chicago, and Salt Lake City, which city is closest to Los Angeles?” We could correctly answer that Salt Lake City is closest. But Salt Lake City is seven hundred miles from Los Angeles, so it is only *relatively* close to Los Angeles—relative to Chicago and New York. Further, even though Salt Lake City is the closest of the candidate set, it is not the closest city of *all cities in the United States of America*. Many cities were not included as candidates—Las Vegas, Tucson, San Diego, and so on. To reliably use a closed-set method such as NSC in stylometry, a researcher must know with reasonable certainty that there are no other possible candidate authors. Without such assurance, the only conclusion that can be drawn is which candidate is the closest from among the set of candidates tested. Because not all possible candidates were included in the Jockers study, statements that make claims about which candidate is the closest of all possible candidates would be unsubstantiated extrapolations and would overstep the

bounds of the evidence. In addition, just because San Diego is close to Los Angeles, that does not mean it is the same as Los Angeles. To claim they are the same city requires more evidence than just a measure of relative proximity. Likewise, in stylometry, relative proximity only connotes similarity of style, not necessarily the same authorship.

Fifth, the NSC probabilities are presented by Jockers et al. as absolute probabilities. This is misleading since, in fact, they are relative probabilities related only to the specific set of candidate authors tested.⁴⁴ Suppose that for some Book of Mormon chapter Rigdon's probability is calculated as 80%, Pratt's probability is calculated as 20%, and each remaining candidate's probability is calculated as nearly 0%. The most that can be concluded from these numbers is that Rigdon's probability of a matching style is four times greater than Pratt's. One could say that the odds are "four to one" (4:1) in favor of Rigdon over Pratt, but one could not meaningfully state Rigdon's calculated likelihood without a comparison to Pratt's. While in a relative sense the probability calculated for Rigdon might be 80% within a limited set of authors, in an absolute sense it might be only 8%, for example, if *all* possible authors were included.

Sixth, the NSC procedure assumes that the variation of the word frequencies in the text blocks is the same for all text blocks. This requirement of equal variance—called homogeneity of variances—is grossly violated in the Jockers study due to the highly disparate sizes of the text blocks. It is completely unreasonable to assume that the variances of word frequencies in text blocks of 100 words are the same as the variances of word frequencies in text blocks of 5,000 words or 15,000 words. Hence the authorship probabilities calculated by NSC make even less sense.

Seventh, the authorship probabilities have still less meaning individually since so many texts (239 chapters) are classified simultaneously in a single statistical procedure. When making a multitude of comparisons within a single test procedure, some of the calculated probabilities will appear to indicate items that are significantly different from each other even though their difference occurred simply by chance. These differences can be spurious and signify nothing. This is a well-known hazard in statistical practice and is referred to as the multiplicity problem.⁴⁵

Naïve or inexperienced analysts frequently make the mistake of overlooking the effects of multiplicity—that is, claiming that a random event has meaning when in reality it is just the result of normal variation in a process.⁴⁶

Eighth, Jockers et al. represented Rigdon's writing style using fourteen articles published in newspapers between 1833 and 1835, as well as nine revelations authored by Rigdon beginning in 1863. The problem is that the styles of these two sets of writings show evidence of being distinctly different, as shown in figure 6, which is based on Jockers et al.'s data.

To confirm this observation, we took all newspaper articles and pamphlets known to have been authored by Sidney Rigdon between 1831 and 1846 to create twenty-five composite texts ranging in size from 2,214 to 8,747 words. We also created fifteen composite texts ranging in size from 3,678 to 6,784 words from all of the sections authored by Sidney Rigdon or jointly by Sidney and Phebe Rigdon in the *Book of the Revelations of Jesus Christ to the Children of Zion through Sidney Rigdon, Prophet, Seer and Revelator*.⁴⁷ The texts were combined in chronological



Sidney Rigdon.

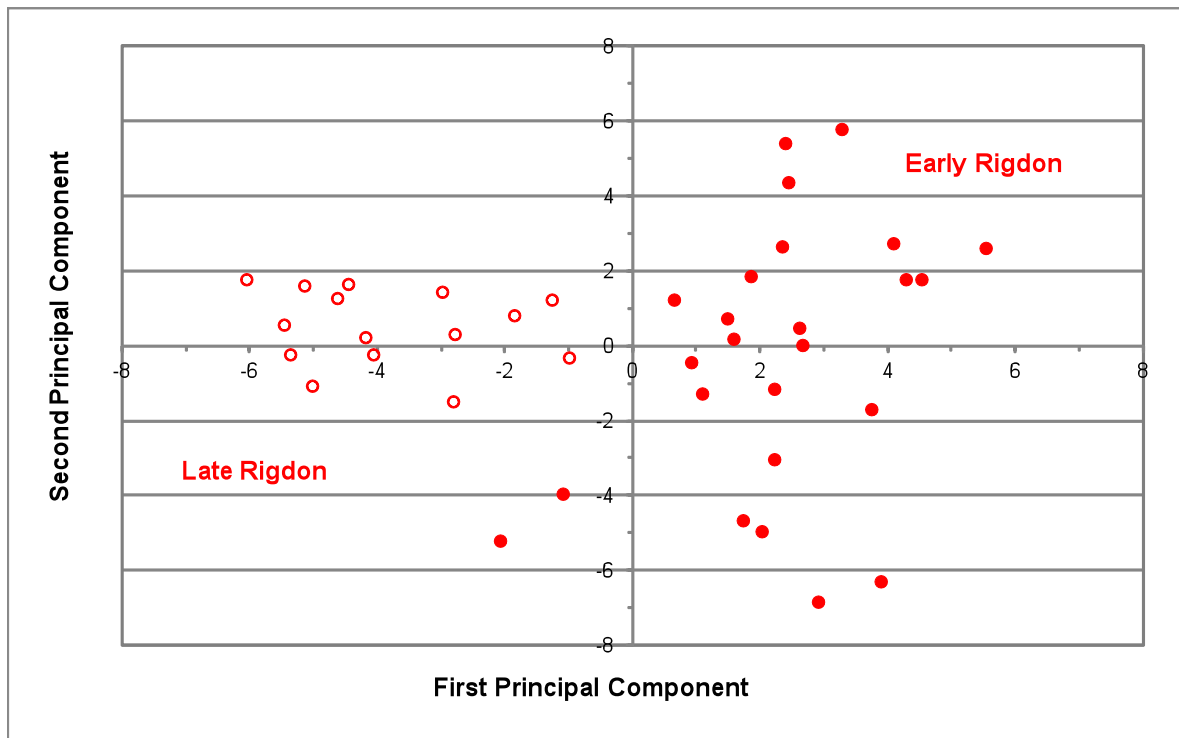


Figure 6. Principal components analysis plot of early and late Rigdon texts. The early Rigdon texts were written from 1833 to 1835 and are shown as solid red dots. The late Rigdon texts were written after 1863 and are shown as open red dots. The distinctness of the two clusters suggests strongly that Rigdon’s early writing style had evolved into another style later in his life.

order, and no section was split between two text blocks. Figure 6 shows the distinct difference in style between the two sets of texts. It is unknown whether Rigdon’s style actually changed over the seventeen intervening years, or whether his revelations reflect the contributions of others such as his wife. In any case, in a study of Book of Mormon authorship, Rigdon’s style should be characterized only by documents written in his early style—the time period closest to the publication of the Book of Mormon. The Rigdon texts used in the Jockers study confound the two Rigdon styles.

The Jockers study concluded:

Our analysis supports the theory that the Book of Mormon was written by multiple nineteenth-century authors, and more specifically, we find strong support for the Spalding-Rigdon theory of authorship. In all the data, we find Rigdon as a unifying force. His signal dominates the book, and where other candidates are more probable, Rigdon is often hiding in the shadows.⁴⁸

In actual fact, the Jockers study has shown nothing. The study design was biased to produce a desired result; the closed-set classification methodology is

completely unsuitable for inferring authorship of the Book of Mormon; the full results for the control texts were ignored; the calculated probabilities were misinterpreted; the chapter-by-chapter probabilities of authorship are not even useful as relative probabilities; the effect of hugely different sample sizes was disregarded; the multiplicity effect of multiple simultaneous testing was ignored; and, finally, Rigdon’s two differing writing styles were confounded into one composite style.

The only idea in the Jockers study that is of some value regarding the Book of Mormon is actually not in their paper, but is based on data listed on their website that we used to produce figure 6. However, it points to a very different conclusion from that drawn by Jockers et al.

Most Recent Study Using ENSC

In response to the Jockers study, we recently conducted a new study correcting the methodological flaws in the Jockers study.⁴⁹ Most important, we developed a modification to the closed-set nearest shrunken centroid (NSC) classification method to enable it to be applied to open-set classification

problems.⁵⁰ We refer to this method as extended nearest shrunken centroid (ENSC) classification. In doing so, we modified the NSC formulas to allow for some other author—that is, to allow for the possibility that an excluded author might have written the text whose authorship is in question. This open-set modification allows for the existence of an unidentified author with writing characteristics nominally consistent with the test text and incorporates this possibility into the probability calculations. Without including the possibility of someone else as the author, if the candidate set does not include the true author (using a closed-set approach for an open-set situation), the probability of similar writing style can be grossly overstated and lead to entirely erroneous interpretations.⁵¹

For purposes of comparability with the Jockers study, we used the same list of 110 characteristic words as Jockers et al. as well as their chapter-by-chapter designation of text blocks from the Book of Mormon. We first reproduced the Jockers study results using the same set of candidate authors to confirm that our implementation of NSC was consistent with theirs. We then repeated the NSC analysis including Joseph Smith in the set of candidate authors. Finally, we applied the open-set ENSC technique allowing for the possibility of some other author. In addition, when we used the ENSC method, we took into account differences in sample sizes, adjusted for multiplicity, and recognized the distinction between Rigdon's time-separated writing styles. Figure 7a displays the results of applying NSC per the Jockers study and applying NSC with Joseph Smith included but without the possibility of someone else as the author. Figure 7b displays the results of applying ENSC allowing for the possibility of some other author.

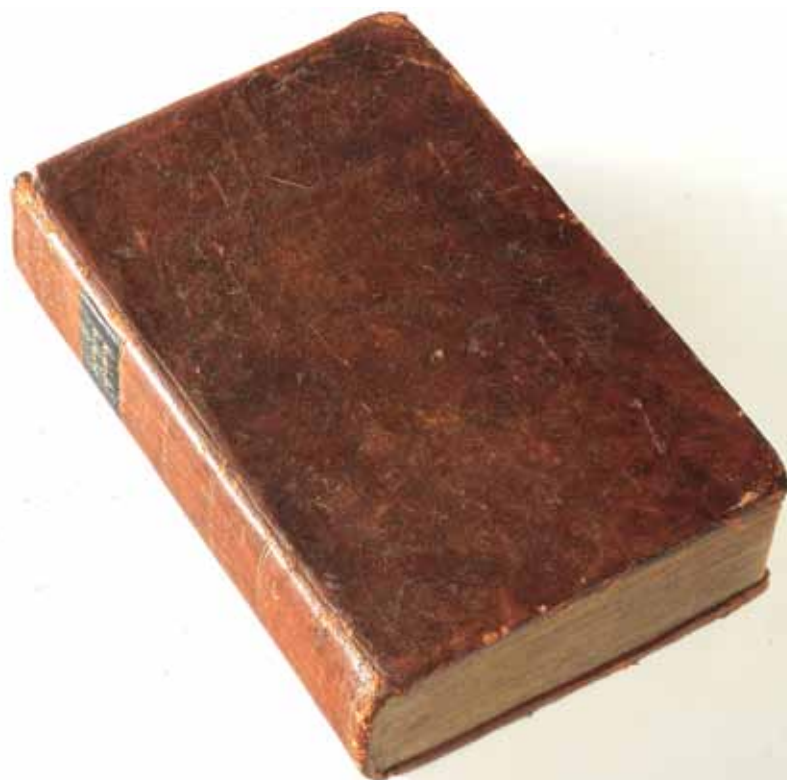
First, examining the NSC graph in figure 7a, we notice that the percentage of chapters NSC assigned to Rigdon is about the same with or without Joseph Smith in the candidate set (39% and 40%, respectively), while the ENSC graph in figure 7b shows far fewer chapters assigned to Rigdon (7%). Interestingly, the ENSC percentage for Rigdon is the sum of roughly equal percentages for early and late Rigdon sample texts.

Next we notice that the percentage for Isaiah/Malachi (26%) as assigned by NSC (fig. 7a) is obviously much too large (as discussed earlier), and the misattribution to Isaiah/Malachi actually increased when Joseph was a candidate author since then NSC

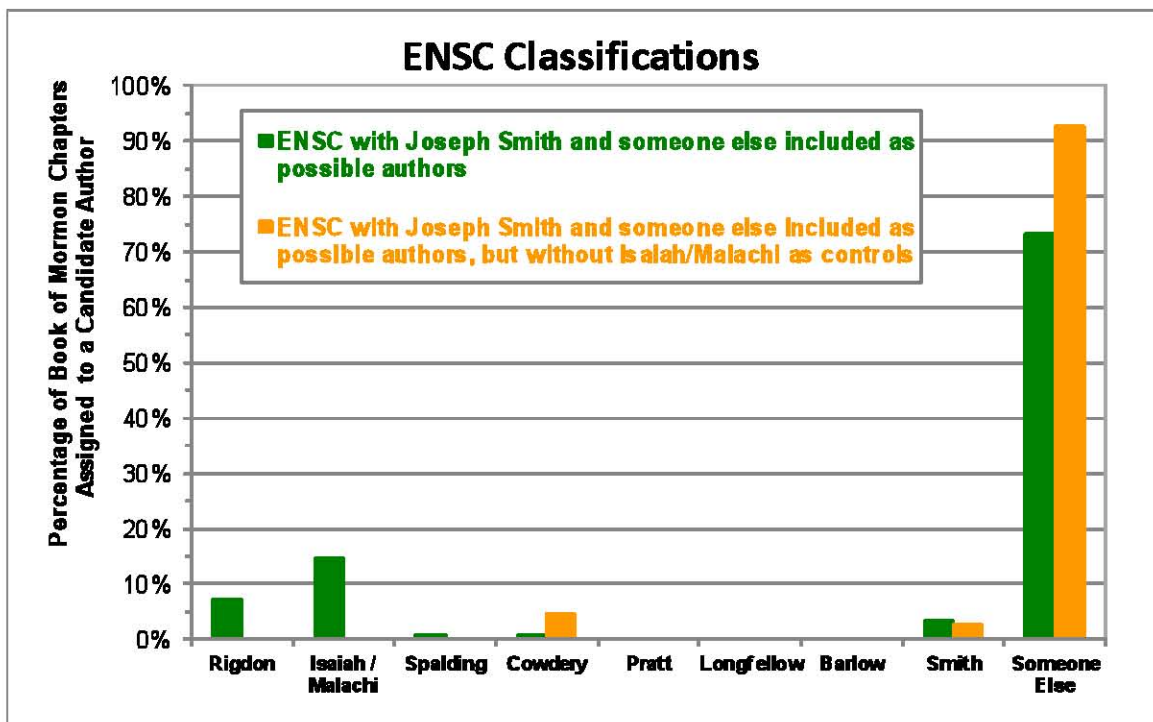
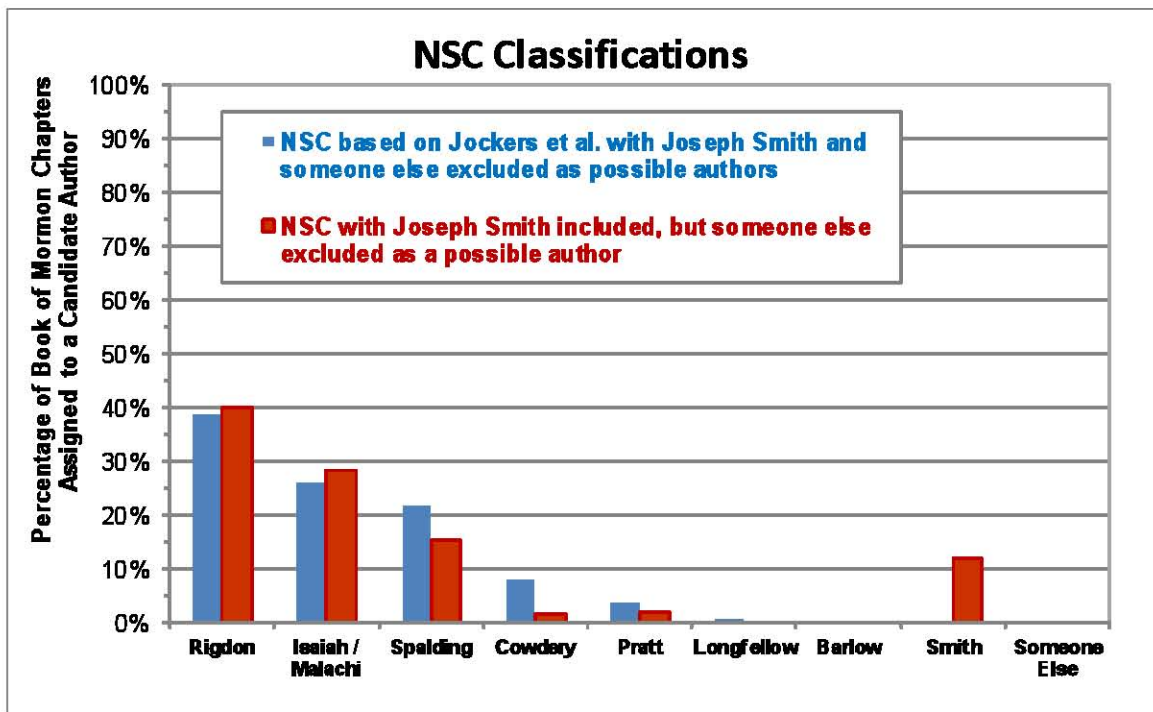
assigned 28% of the chapters to Isaiah/Malachi. However, the ENSC-assigned percentage of 15% (fig. 7b) is much closer to the correct percentage (12%).

Considering Spalding without Joseph Smith as a candidate author, NSC assigned 22% of the chapters to Spalding, yet only 15% to him when Joseph Smith was included (fig. 7a). Obviously, when Joseph Smith is included in the analysis, any supposed support for the Spalding-Rigdon theory diminishes. With Joseph Smith in the candidate author set, we see that NSC assigned 12% of the chapters to Joseph Smith because of chapter reassignment away from Spalding, Cowdery, and Pratt. This seems consistent with the claim that Joseph Smith, as translator, dictated the text of the Book of Mormon, and in doing so perhaps had some influence on the structure of language in the document.

In contrast, when applying ENSC, the combined total for Spalding-Rigdon drops to only 8%, with ENSC assigning a mere 3% to Joseph Smith (fig. 7b). The few chapters that ENSC indicated to be closest to Rigdon, Spalding, Cowdery, and Smith are randomly dispersed throughout the 239 chapters,



Book of Mormon, 1st edition (1830). © 1995 IRI.



Figures 7a and 7b. Nearest shrunken centroid (NSC) and extended nearest shrunken centroid (ENSC) classification methods applied to Book of Mormon authorship. Although the closed-set NSC technique assigns a majority of chapters to Spalding and Rigdon within a constrained set of candidate authors, when allowing for the possibility that the candidate set is incomplete, the open-set ENSC technique assigns an even larger majority of the chapters to an unidentified author who was not included in the NSC candidate set. Percentages are based on the number of chapters that are deemed closest to a candidate author's style.

indicating that they should be considered random misclassifications.

Most interesting, though, the ENSC method (fig. 7b) assigned 73% of the chapters to “Someone Else.” Further, excluding the Isaiah/Malachi chapters and looking only at the non-Isaiah/Malachi Book of Mormon chapters, ENSC assigned 93% of those chapters to “Someone Else” with a few chapters randomly assigned to Oliver Cowdery and Joseph Smith, as would be expected if Joseph Smith had translated the text with Oliver Cowdery as his scribe.

Clearly Jockers et al.’s claim of astronomical probabilities in support of the Spalding-Rigdon theory is a great exaggeration. The ENSC results confirm our analysis that the Jockers study was fatally flawed in concept and execution. Contrary to their contention, the evidence does not provide credible support for the claim that the writing styles exhibited in the Book of Mormon match any of their candidate authors—Spalding, Rigdon, Cowdery, or Pratt. In fact, the evidence from a correctly conducted analysis clearly supports the claim that *someone other than their set of candidate authors wrote the book*. Therefore, based on these findings, we conclude that stylometric evidence does not support the Spalding-Rigdon theory of Book of Mormon authorship.

Conclusion

Stylometric analyses of the Book of Mormon have generated much interest over the past thirty years. Some of these analyses have produced interesting information, but some of the studies have been characterized by hyperbole, faulty reasoning, and misapplication of statistical methods. When ex-

amining all the evidence, our overall conclusion is that the Book of Mormon displays multiple writing styles throughout the text consistent with the book’s claim of multiple authors and that the evidence does not show the writing styles of alleged nineteenth-century authors to be similar to those in the Book of Mormon. Further, the claims thus far put forward for alternative authorship of the Book of Mormon, other than as described by Joseph Smith, are untenable. ■



Matthew Roper (MA, Brigham Young University) is a research scholar for the Neal A. Maxwell Institute for Religious Scholarship, Brigham Young University.



Paul J. Fields (PhD, Pennsylvania State University) is a consultant specializing in research methods and statistical analysis. He has extensive experience in textual analysis and linguistic computing.



G. Bruce Schaalje (PhD, North Carolina State University) is a professor of statistics at Brigham Young University.

NOTES

1. Louis C. Midgley, “Who Really Wrote the Book of Mormon? The Critics and Their Theories,” in *Book of Mormon Authorship Revisited: The Evidence for Ancient Origins*, ed. Noel B. Reynolds (Provo, UT: FARMS, 1997), 101-39.
2. Eber D. Howe, *Mormonism Unveiled, or, a Faithful Account of That Singular Imposition and Delusion, from Its Rise to the Present Time* (Painesville, OH: Printed and Published by the Author, 1834).
3. Matthew Roper, “The Mythical ‘Manuscript Found,’” *FARMS Review* 17/2 (2005): 7-140; and Roper, “Myth, Memory, and ‘Manuscript Found,’” *FARMS Review* 21/2 (2009): 179-223.
4. John F. Burrows, “Computers and the Study of Literature,” in *Computers and Written Texts*, ed. Christopher S. Butler (Oxford: Blackwell, 1992), 167-204.
5. Carnes Lord, “On the Early History of the Aristotelian Corpus,” *American Journal of Philology* 107/2 (1986): 137-61.
6. Reginald C. Churchill, *Shakespeare and His Betters: A History and a Criticism of the Attempts Which Have Been Made to Prove That Shakespeare’s Works Were Written by Others* (Bloomington: Indiana University Press, 1958); James G. McManaway, *The Authorship of Shakespeare* (Washington, DC: Folger Shakespeare Library, 1962); and Hugh Craig and Arthur F. Kinney, *Shakespeare, Computers, and the Mystery of Authorship* (Cambridge: Cambridge University Press, 2009).
7. J. D. James, *The Genuineness and Authorship of the Pastoral Epistles* (London: Longmans, Green, 1906); Percy N. Harrison, *The Problem of the Pastoral Epistles* (London: Oxford University Press, 1921).
8. Joseph Black, “The Rhetoric of

- Reaction: The Martin Marprelate Tracts (1588–89), Anti-Martinism, and the Uses of Print in Early Modern England,” *Sixteenth Century Journal* 28/3 (1997): 707–25.
9. See Frederick Mosteller and David L. Wallace, *Inference and Disputed Authorship: “The Federalist”* (Reading, MA: Addison-Wesley, 1964).
 10. David I. Holmes, “The Evolution of Stylometry in Humanities Scholarship,” *Literary and Linguistic Computing* 13/3 (1998): 112.
 11. T. C. Mendenhall, “The Characteristic Curves of Composition,” *Science* 214 (11 March 1887): 237–46.
 12. C. Mascol, “Curves of Pauline and Pseudo-Pauline Style I,” *Unitarian Review* 30 (November 1888): 452–60; Mascol, “Curves of Pauline and Pseudo-Pauline Style II,” *Unitarian Review* 30 (December 1888): 539–46.
 13. L. A. Sherman, *Analytics of Literature: A Manual for the Objective Study of English Prose and Poetry* (Boston: Ginn, 1893).
 14. Mosteller and Wallace, *Inference and Disputed Authorship*.
 15. John F. Burrows, “Word Patterns and Story Shapes: The Statistical Analysis of Narrative Style,” *Literary and Linguistic Computing* 2/1 (1987): 61–70; Burrows, “‘An Ocean Where Each Kind . . .’: Statistical Analysis and Some Major Determinants of Literary Style,” *Computers and the Humanities* 23 (1989): 309–21; Burrows, “‘Delta’: A Measure of Stylistic Difference and a Guide to Likely Authorship,” *Literary and Linguistic Computing* 17/3 (2002): 267–87; Burrows, “Questions of Authorship: Attribution and Beyond,” *Computers and the Humanities* 37 (2003): 5–32.
 16. Principal components analysis is a multivariate statistical technique that can reduce a large set of correlated variables to a smaller set of uncorrelated variables that are linear combinations of the original variables and are arranged in order of decreasing importance.
 17. While fingerprints and DNA profiles can be precisely measured, an author’s word-use preferences are far more variable and nebulous. For example, when counting how frequently an author uses the word *the* in multiple samples of the author’s writing, the counts form a cluster rather than a single dot, since an author does not use *the* with exactly the same frequency in every text. In contrast, a person’s fingerprints and DNA profile do not vary.
 18. Harold Love, *Attributing Authorship: An Introduction* (Cambridge: Cambridge University Press, 2002).
 19. Wayne A. Larsen, Alvin C. Rencher, and Tim Layton, “Who Wrote the Book of Mormon? An Analysis of Wordprints,” *BYU Studies* 20/3 (1980): 225–51; reprinted by Wayne A. Larsen and Alvin C. Rencher in *Book of Mormon Authorship: New Light on Ancient Origins*, ed. Noel B. Reynolds (Provo, UT: BYU Religious Studies Center, 1982), 157–88; John L. Hilton, “On Verifying Wordprint Studies: Book of Mormon Authorship,” *BYU Studies* 30/3 (1990): 89–108; reprinted in Reynolds, *Book of Mormon Authorship Revisited*, 225–53.
 20. David I. Holmes, “A Stylometric Analysis of Mormon Scripture and Related Texts,” *Journal of the Royal Statistical Society A* 155, part 1 (1992): 91–120.
 21. Matthew L. Jockers, Daniela M. Witten, and Craig S. Criddle, “Reassessing Authorship of the Book of Mormon Using Delta and Nearest Shrunken Centroid Classification,” *Literary and Linguistic Computing* 23/4 (2008): 465–91.
 22. Discriminant analysis is a statistical procedure that uses multiple variables (in this case, word frequencies) to classify new observations (text block of unknown authorship) using mathematical functions that weight each of the variables to maximize the differences between known groups (text blocks of known authorship) while minimizing the differences within the groups. Thus, dissimilar groups can be distinguished (discriminated) from similar groups.
 23. Larsen, Rencher, and Layton, “Who Wrote the Book of Mormon?,” 240.
 24. Larsen, Rencher, and Layton, “Who Wrote the Book of Mormon?,” 245. The reprinted article toned down the claim to read, “The evidence to date is that many authors wrote the Book of Mormon.” Larsen and Rencher, “Who Wrote the Book of Mormon?,” 180.
 25. Hilton, “On Verifying Wordprint Studies,” 89–108.
 26. Andrew Q. Morton, *Literary Detection: How to Prove Authorship and Fraud in Literature and Documents* (New York: Scribner’s Sons, 1978).
 27. This has now been published as Royal Skousen, ed., *The Printer’s Manuscript of the Book of Mormon: Typographical Facsimile of the Entire Text in Two Parts* (Provo, UT: FARMS, 2001).
 28. Hilton, “On Verifying Wordprint Studies,” in Reynolds, *Book of Mormon Authorship Revisited*, 241.
 29. Todd K. Moon, Peg Howland, and Jacob H. Gunther, “Document Author Classification Using Generalized Discriminant Analysis,” in *Proceedings of the Fourth Workshop on Text Mining, Sixth SIAM International Conference on Data Mining* (paper presented at the 2006 SIAM Conference on Data Mining, Bethesda, Maryland, 20–22 April 2006).
 30. As a side note, we mention another Book of Mormon stylometry study reported in 1984 by psychiatrist Ernest H. Taves: *Trouble Enough: Joseph Smith and the Book of Mormon* (Buffalo, NY: Prometheus Books, 1984). In his Book of Mormon study Taves used word-pattern ratios similar to the Hilton study but concluded that “the texts provide no evidence for multiple authorship” in the Book of Mormon. In a thorough and systematic review of Taves’s study, Hilton showed that the study was fatally flawed from improper text sampling, misapplication of the chi-square test statistic, and other design weaknesses. John L. Hilton, review of *Book of Mormon Stylometry*, by Ernest Taves (FARMS Preliminary Report, 1986), 16; in light of the amateurish nature of Taves’s study, Hilton advised—and we agree—that its usefulness is limited to showing how stylometry “should not be done.” See Kenneth H. Godfrey, “Not Enough Trouble,” review of *Trouble Enough: Joseph Smith and the Book of Mormon*, by Ernest Taves, *Dialogue* 19/3 (1986): 139–44.
 31. David I. Holmes, “Vocabulary Richness and the Prophetic Voice,” *Literary and Linguistic Computing* 6/4 (1991): 259–68.
 32. Holmes, “Stylometric Analysis of Mormon Scripture,” 91.
 33. Holmes, “Stylometric Analysis of Mormon Scripture,” 117, 118.
 34. This was pointed out to Holmes in a letter to the editor saying, “In general, conclusions about statistical similarity require knowledge of the power of the statistical procedure to detect differences.” G. Bruce Schaalle, letter to the editors, *Journal of the Royal Statistical Society A* 156, part 1 (1993): 115. Holmes replied, “My research showed that the multivariate ap-

- proach to measuring vocabulary richness successfully discriminates between samples from within the same genre. Both the personal writings and the prophetic voice of Joseph Smith differ from those of Joanna Southcott in the element of style represented by vocabulary richness.” David I. Holmes, author’s reply, *Journal of the Royal Statistical Society A* 156, part 1 (1993): 116. Apparently Holmes was unwilling to acknowledge that one successful application of a method in a specific situation does not prove the method to be generally powerful in all situations. See Holmes, “Vocabulary Richness and the Prophetic Voice,” 259–68.
35. G. Bruce Schaalje, John L. Hilton, and John B. Archer, “Comparative Power of Three Author-Attribution Techniques for Differentiating Authors,” *Journal of Book of Mormon Studies* 6/1 (1997): 47–63. Comparing the discriminating power of Holmes’s method to other methods when applied to texts of undisputed authorship, the discriminating power of vocabulary richness measures is much lower than that of noncontextual word frequencies or word-pattern ratios. R. Harald Baayen, *Word Frequency Distributions* (Boston: Kluwer Academic, 2001), 214.
 36. David I. Holmes and R. S. Forsyth, “The *Federalist* Revisited: New Directions in Authorship Attribution,” *Literary and Linguistic Computing* 10/2 (1995): 111–27.
 37. David L. Hoover, “Another Perspective on Vocabulary Richness,” *Computers and Humanities* 37 (2003): 173.
 38. Jockers, Witten, and Criddle, “Reassessing Authorship of the Book of Mormon,” 465–91.
 39. Robert Tibshirani, Trevor Hastie, Balasubramanian Narasimhan, and Gilbert Chu, “Class Prediction by Nearest Shrunken Centroids, with Applications to DNA Microarrays,” *Statistical Science* 18/1 (2003): 104–17.
 40. *Personal Writings of Joseph Smith*, comp. and ed. Dean C. Jessee, rev. ed. (Salt Lake City: Deseret Book, 2002). Holographic material of Joseph Smith is highlighted in bold in this collection.
 41. See supplementary data for Jockers, Witten, and Criddle, “Reassessing Authorship of the Book of Mormon,” 465–91, at <http://purl.stanford.edu/ir:rs276tc2764> (accessed 29 February 2012).
 42. Jockers, Witten, and Criddle, “Reassessing Authorship of the Book of Mormon,” 472.
 43. Jockers, Witten, and Criddle, “Reassessing Authorship of the Book of Mormon,” 473, table 1.
 44. A probability can be absolute only if it includes all possible outcomes. The probability of rolling a 2 on a six-sided die is 1 out of 6 (approximately 17%). This is an absolute probability since the event of rolling a 2 is one of six possible outcomes among the complete set of events—rolling a 1, 2, 3, 4, 5, or 6. But when considering the events of rolling a 1 or rolling a 2 as the only admissible outcomes and thereby ignoring the events of rolling a 3, 4, 5, or 6, the probability of rolling a 2 is 1 out of 2 (50%). While the absolute probability of rolling a 2 among all possible events is 17%, the relative probability of rolling a 2 among the limited set of events of rolling a 1 or rolling a 2 is 50%. Confusing an event’s relative probability with its absolute probability can greatly exaggerate the perceived likelihood of that event.
 45. When making multiple comparisons simultaneously, a researcher can make an error in statistical inference by claiming two things are significantly different when in fact they are not. When viewed individually, a pair of items may appear to be different, yet when viewed in the full context of all possible comparisons in the study, their difference can be negligible. The probability of making such an error increases as the number of multiple comparisons increases. Consequently, the more complex a problem is, the greater the potential to make such an error. Standard statistical techniques have been developed and are used by competent researchers to compensate for the effects of multiplicity and to guard against making such inferential errors. The multiplicity problem is a major issue to guard against in genetic association studies wherein millions of genetic markers can be analyzed simultaneously. Researchers in one study may report statistically significant results, yet subsequent researchers will not be able to reproduce those results. One of the possible reasons for such conflicting outcomes is the failure of the first researchers to account for multiplicity. Since NSC is a classification procedure originally intended for use in genomics applications, the failure of Jockers et al. to recognize the potential for multiplicity to produce misleading results indicates that the researchers were unfamiliar with the tool (NSC) they were using and unskilled in its proper use.
 46. Please note that the presence of a few extreme values—either high or low—would be expected in any data set, even if the data come from a process that is truly random. Their presence does not necessarily signify anything unusual in the data. Researchers who ignore the multiplicity problem are prone to finding evidence in the data that supports their preconceived notion of what “should be” in the data. This is often referred to as data snooping, in contrast to data analysis. Such researchers are determined to find in the data what they want to be in the data regardless of facts and reason.
 47. Located in Stephen Post Papers, folders 11 and 12, L. Tom Perry Special Collections, Harold B. Lee Library, Brigham Young University, Provo, Utah.
 48. Jockers, Witten, and Criddle, “Reassessing Authorship of the Book of Mormon,” 483.
 49. G. Bruce Schaalje, Paul J. Fields, Matthew Roper, and Gregory L. Snow, “Extended Nearest Shrunken Centroid Classification: A New Method for Open-Set Authorship Attribution of Texts of Varying Sizes,” *Literary and Linguistic Computing* 26/1 (2011): 71–88.
 50. G. Bruce Schaalje and Paul J. Fields, “Open-Set Nearest Shrunken Centroid Classification,” *Communications in Statistics: Theory and Methods* 41 (2012): 638–52.
 51. The fundamental fallacy in the Jockers study was that they equated genomics problems to stylometry problems. Although NSC has proven to be highly successful in genomics classification problems, stylometric problems are much different: a large set of texts is usually the subject of classification, the sample sizes vary over a wide range, and most important, the set of candidate authors usually cannot be assumed to be closed. These characteristics are not present in the typical genomics analysis. Consequently, naïve application of NSC, as in the Jockers study, can produce highly misleading results. A reanalysis of the Book of Mormon using ENSC produced dramatically different results from the NSC method.