

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/250900139>

Reassessing authorship of the Book of Mormon using delta and nearest shrunken centroid classification

Article *in* Literary and Linguistic Computing · September 2008

DOI: 10.1093/lc/fqn040

CITATIONS

58

READS

2,113

3 authors, including:



Matthew L. Jockers

Apple Inc.

69 PUBLICATIONS 832 CITATIONS

[SEE PROFILE](#)



Craig S. Criddle

Stanford University

336 PUBLICATIONS 15,219 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Bioelectrochemical Systems [View project](#)



Reclaimed water [View project](#)

NOTE: THIS IS A PRE-PRINT DRAFT VERSION OF: "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification" published in *Literary and Linguistic Computing*. The published version includes a variety of editorial changes, several corrections of fact, and one correction of a data processing error. Interested readers are advised to consult the published version of this paper:

Jockers, Matthew L., 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*, December 2008; 23: 465 - 491; doi 10.1093/lc/fqn040 (online at <http://lc.oxfordjournals.org/cgi/content/short/fqn040v2>)

Title: Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification

Authors: Matthew L. Jockers¹, Daniela M. Witten, and Craig S. Criddle

Abstract: Mormon prophet Joseph Smith (1805-1844) claimed that more than two-dozen ancient individuals (Nephi, Mormon, Alma, *etc.*) living from around 2200 BC to 421 AD authored the *Book of Mormon* (1830), and that he translated their inscriptions into English. Later researchers who analyzed selections from the *Book of Mormon* concluded that differences between selections supported Smith's claim of multiple authorship and ancient origins. We offer a new approach that employs two classification techniques: "delta" commonly used to determine probable authorship and "nearest shrunken centroid" (NSC), a more generally applicable classifier. We use both methods to determine, on a chapter-by-chapter basis, the probability that each of seven potential authors wrote or contributed to the *Book of Mormon*. Five of the seven have known or alleged connections to the *Book of Mormon*, two do not, and were added as controls based on their thematic, linguistic, and historical similarity to the *Book of Mormon*. Our results indicate that likely nineteenth century contributors were Solomon Spalding, a writer of historical fantasies; Sidney Rigdon, an eloquent but perhaps unstable preacher; and Oliver Cowdery, a schoolteacher with editing experience. Our findings support the hypothesis that Rigdon was the main architect of the *Book of Mormon* and are consistent with historical evidence

suggesting that he fabricated the book by adding theology to the unpublished writings of Spalding (then deceased).

Background

Since its publication in March 1830, the origin of the *Book of Mormon*—particularly its claim of ancient origins—has been the subject of intense scrutiny and debate. Mormon prophet Joseph Smith Jr. (1805-1844) claimed that an angelic messenger delivered to him a record written from around 2200 BC to 421 AD by ancient Native Americans in “reformed Egyptian.” Smith claimed to have used a seer stone to translate the record into English. By February 1831, two competing theories had appeared: Alexander Campbell (1831), founder of the Campbellite religious movement, proposed Smith himself as author while the *Cleveland Advertiser* (1831) proposed Sidney Rigdon.² Rigdon was a former Campbellite preacher who acquired ecclesiastical status on par with Smith’s almost immediately after his rapid conversion in October 1830. Campbell eventually (1844) concluded that Rigdon was the probable author of the *Book of Mormon*, but his initial explanation held that Smith wrote the book by drawing from sermons and local folklore.³

One year later, in February 1832, a third candidate-author was named when Mormon missionaries Orson Hyde and Samuel Smith read passages from the *Book of Mormon* at a schoolhouse in Conneaut (New Salem), Ohio. Nehemiah King, who was present at these readings, claimed that Hyde “had preached from the [novelistic] writings of Solomon Spalding” (Wright, 1833).⁴

Spalding (often "Spaulding") was a frustrated novelist who, prior to his death in 1816 (i.e. from 1811 to 1815), shared his unpublished novel with his neighbors, family, and associates in Conneaut. In 1833 the Spalding allegations came to the attention of E. D. Howe, who joined with ex-Mormon Philastus Hurlbut to investigate the matter. Hurlbut collected affidavits from Spalding's former neighbors and family in Conneaut⁵. The witnesses recalled having heard much of the plot and several names from the *Book of Mormon* in a draft novel titled "Manuscript Found," a now-lost text that Spalding submitted for publication to a Pittsburgh publisher in late 1812. In *Mormonism Unveiled* [sic] (1834, 1977) Howe linked Rigdon to Spalding through this publisher. The resulting "Spalding-Rigdon Theory" holds that Rigdon acquired the Spalding manuscript through his connections to the Pittsburgh publishing shop, added his own theology, and then revealed it to the public through Smith as the *Book of Mormon*. In the years after Howe's publication, others provided testimony supportive of the theory, including Spalding's widow, Spalding's daughter, the owner of the Pittsburgh publishing shop, and others who claimed that Spalding shared his work with them⁶ or who claimed to have seen a copy of "Manuscript Found" after Spalding's death.⁷ Expansions to the theory followed, including a detailed analysis of Rigdon's life by William H. Whittsitt (1886, 1891), and of other likely collaborators, including Smith's second cousin Oliver Cowdery, a schoolteacher with editing experience (Deming, 1888) and Parley P. Pratt, a former disciple of Rigdon (Schroeder, 1901).⁸

Throughout the nineteenth and early twentieth century, the Spalding-Rigdon Theory was the favored explanation for the origin of the *Book of Mormon*, but Fawn Brodie's rejection of this theory in her controversial biography of Smith (1945) marked a turning point in the debate. Invoking witness tampering and "false memory syndrome," Brodie dismissed the affidavits collected by Hurlbut. She believed that a Spalding holograph discovered in Honolulu, Hawaii, and stored within a large envelope with the penciled-in title "Manuscript Story—Conneaut Creek," (1810) was in fact the lost Spalding document known to the Conneaut witnesses as "Manuscript Found."⁹ Despite having no evidence that the Honolulu manuscript was the same text that the Conneaut witnesses heard Spalding read to them (and subsequently recognized as a source text for the *Book of Mormon*), Brodie nonetheless concluded that Spalding could not have been an author of the *Book of Mormon* because the similarities between the *Book of Mormon* and the text found in Hawaii were "not sufficient to justify the thesis of common authorship."¹⁰ Her rejection of the Spalding-Rigdon Theory was so widely accepted that the Spalding-Rigdon Theory came to be regarded by most students of Mormon history as "an historiographical artifact without credibility among serious scholars" (Bushman, 2005).

Among contemporary secular scholars of Mormonism, the theory of Smith as solitary author is a generally accepted explanation. Twentieth century advocacy of this theory began with I. Woodbridge Riley (1902) who proposed that Smith drew inspiration from locally available source materials, including Ethan Smith's *View of the Hebrews* (1825). Riley also speculated that Joseph

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

Smith's writing was influenced by epilepsy-induced visions and that Smith created characters modeled on members of his family, including himself. Brigham Roberts (1857-1933), a Mormon leader and intellectual whose writings are collected in *Studies of the Book of Mormon* (Madsen, 1985), likewise concluded that Smith had the imagination and source material to produce the *Book of Mormon* on his own (1985). Smith's textual sources, Roberts argued, likely included *View of the Hebrews* (1825) and Josiah Priest's *Wonders of Nature and Providence* (1825). Brodie advanced similar arguments (1945), and followed Riley's footsteps with speculation regarding Smith's psychology (1971). In recent work, David Persuitte (2000) provides textual parallels to strengthen connections to both Ethan Smith and Josiah Priest, while historian Dan Vogel (2004) expands the psychological speculation, suggesting that the *Book of Mormon* is best explained as the result of Smith family dynamics and Smith's willingness to engage in a pious fraud.

In addition to historical studies of Smith and the origins of the *Book of Mormon* (such as those noted above), there have been a smaller number of quantitative, or "stylometric," studies. A team of Brigham Young University researchers led by Wayne Larsen conducted the first among these (Larsen 1980).¹¹ Employing multivariate, cluster, and classification analysis, Larsen, Rencher, and Layton set out to test the proposition that the *Book of Mormon* is the work of a single author (perhaps Smith) or multiple authors (ancient or nineteenth century). Larsen's study included analysis of thirty-eight frequently occurring common non-contextual words and forty-two rarely occurring non-

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

contextual words. To generate frequency lists, the researchers first assumed that "the writers of each verse, or partial verse, could be identified according to information given in the text" and thus they "assigned" verses and partial verses to classes based on their "careful scrutiny" of the text (1980). They concluded from statistical analysis of this material that the text was not the work of Joseph Smith and that many authors likely wrote it. Using samples of known writings from Solomon Spalding, Sidney Rigdon, and other Smith contemporaries, Larsen claimed further that the multiple styles they detected in the *Book of Mormon* were not likely to be the work of any of these 19th century authors.

Several problems are now apparent in the methodologies employed by Larsen *et. al.* First, they grouped verses and partial verses from the *Book of Mormon* into clusters based on their understanding of speakers (or characters) in the *Book of Mormon* (i.e. Nephi, Alma, *etc.*). Because the characters had distinctive vocabulary "wordprints" within these selections, they concluded that the *Book of Mormon* was a multi-authored work.¹² They further reasoned that because their selections did not match the styles of potential nineteenth-century authors, they could conclude that the text was not the work of a nineteenth-century author. However, their analysis did not exclude the possibility that their chosen selections were composites containing different fractional contributions from different nineteenth-century authors.

A further problem stems from Larsen's reliance upon context sensitive words. Though Larsen claims to use only non-contextual words, his list of selected words is questionable. It includes words such as "behold," "forth," "lest,"

"nay", "O," "unto," "wherefore," and "yea"—words that are common in scripture and thus contextual. They occur at a much higher frequency in the *Book of Mormon* than in the writings of nineteenth-century authors. Take the word "unto," for example: It occurs 3610 times in the *Book of Mormon*, a rate of 135 occurrences for every ten thousand words. In the entire Chadwyck-Healey Early American Fiction collection (2000), a collection of 875 novels spanning the period from 1789 to 1875, the word "unto" appears 2346 times, a rate of just 3.8 occurrences for every 10,000 words¹³. Even sympathetic scholars, such as the statistician D. James Croft (1981), caution against reading too much into Larsen's results.¹⁴

In a paper from around 1988,¹⁵ Mormon investigator John L. Hilton claimed that his group had significantly improved Larsen's techniques and that their results reconfirmed his conclusion that the *Book of Mormon* is a work of multiple, though ancient, authors. For his analysis of the *Book of Mormon*, however, Hilton chose to analyze subjectively grouped and edited selections from the *Book of Mormon* put together in the form of 5000-word blocks of text. Like Larsen, Hilton assumed that characters such as Nephi and Alma can be viewed as candidate authors, and he selected blocks of text from what he referred to as "didactic" sections for the characters "Nephi" and "Alma." He then followed Larsen in assuming that each selection could only be the work of a single 19th century author, not the work of multiple nineteenth century authors. At best, one might hope to conclude from such an analysis that the chosen selections are not by the same author, but the methodology used does not

exclude the possibility of multiple nineteenth century authors. Hilton's methodology thus did not address a key aspect of the *Book of Mormon* authorship question.

In Appendix 3 of his essay, Hilton identifies the sources for his compilation: not a single manuscript, or the published 1830 version of the *Book of Mormon*, but instead, a composite compilation of selections from four sources based upon what he and his team judged to be the oldest. The provenance of this material is questionable. Also problematic is that Hilton's compilation of old Mormon manuscripts did *not* include significant sections and direct quotations from the King James Bible—sections and quotations that are an acknowledged part of the 1830 *Book of Mormon*.¹⁶ Most importantly, Hilton's analysis neglected to include a comparison with the work of Rigdon. This omission is difficult to understand given the other potential authors whose work Hilton analyzed. In our work, we include a large amount of newly available Rigdon text of certain provenance, adding to the limited amount available at the time of Larsen's study.

More compelling than the work of Hilton and Larsen is the work of statistician David Holmes (1985, Holmes, 1991a, Holmes, 1991b, Holmes, 1992). In separate papers from 1991 and 1992, Holmes investigated *Book of Mormon* authorship using a multivariate measurement of vocabulary richness. Holmes compared the *Book of Mormon* to thirteen writing samples from Joseph Smith, Joanna Southcott, and the King James Bible.¹⁷ He measured the richness of noun usage in the various works: a technique that Holmes claims enables him to discriminate between the "personal" and the "prophetic" writings of Joseph Smith

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

as well as between the personal writings of Smith and those of Joanna Southcott. Using this technique, Holmes further discriminates between the prophetic voice of Smith and that of Southcott. Holmes's derives the "signal" for Smith's prophetic voice from Smith's revelations as they are recorded in *Doctrine and Covenants*; the personal voice he derives from the letters and diary entries collected in Dean Jessee's *The Personal Writings of Joseph Smith*.

Detecting differences between Smith's prophetic and personal voice was a key discovery for Holmes. His technique appeared to prove effective in discriminating between authors and between authorial voices in different contexts. From this, Holmes argued that his multivariate measurements of vocabulary richness offered no evidence to support the argument that the *Book of Mormon* is a work of multiple authors. This conclusion stood in direct contradiction to the previous analyses by Larsen and Hilton. However, two problems are apparent in Holmes's work: First, his reliance upon the letters and diary entries collected by Dean C. Jessee in *Personal Writings of Joseph Smith* (Smith and Jessee, 2002) as a reliable source for Smith's personal voice and second his reliance upon the *Doctrine and Covenants* as a reliable source for Smith's prophetic voice.

Though Holmes was careful to select "only those letters written by Smith himself [in Smith's hand], or preserved in the handwriting of clerks who state specifically that Smith is dictating" (Holmes, 1991b) even this subset of Dean Jessee's collection is problematic. In the opening sentence of his introduction to *Personal Writings of Joseph Smith*, Jessee declares: "it matters very little

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

whether or not a person writes his own journals, letters, and speeches or delegates others to write for him" (Jessee, 2002). His point here is that even if written by others, the material reflects the mind of the Smith if not the actual words as written. For authorship attribution analysis, however, we are less concerned with whether a document captures the "spirit" of an attributed author and more specifically interested in whether the document is written by and in the natural style of the attributed author. With Smith, however, we cannot reasonably conclude this point, that the documents attributed to him are indeed reflections of his individual literary style. On the contrary, in studying Smith and reading Jessee's collection of documents, one becomes immediately and acutely aware of how little we can, even blithely, attribute to Smith and Smith alone.

Jessee notes the problems associated with claiming that Smith was the author of the words attributed to him: "His philosophy" writes Jessee, "was that 'a prophet cannot be his own scribe.'"¹⁸ Indeed, even Jessee avoids use of the word "author" preferring instead "writings attributed to him [Smith]." Jessee points out that while Smith "produced a sizable collection of papers, the question remains as to how clearly they reflect his own thoughts and personality [because] we inherit the limitations that produced them . . . the wide use of clerks taking dictation or even being assigned to write for him, and the editorial reworking of reports of what he did and said" (Smith and Jessee, 2002). Jessee notes further that the "practice . . . of inserting eyewitness writings that have been changed from indirect to direct discourse . . . gives the impression that Joseph wrote them," when in fact he did not. Referring to one particular case, Jessee writes

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

that the impressions of Joseph Smith given . . . probably reflect the personality of the editor more than they do Joseph's." Even for the twenty-three letters in Smith's hand, which Jessee republishes in facsimile form, we cannot easily assume that Smith is the sole author. Many of the letters in Jessee's collection show the handwriting of Smith along side and intermingled with the handwriting of other authors, including Rigdon and Cowdery. Even when writing something as personal as a journal entry or letter, we see consistent evidence of collaboration and co-authorship. Unfortunately, such writing cannot be used as a reliable sample of known authorship.¹⁹

Second, and equally problematic, is Holmes's use of the *Doctrine and Covenants* as a reliable example of Smith's prophetic voice. This text of revelations is *ascribed* to Smith, but as is the case with many of his letters and diary entries, he did not write it unaided. Rather, he is reported to have dictated the revelations to one of his scribes. From 1829 to 1838, two of Smith's main scribes were none other than Sidney Rigdon and Oliver Cowdery, who according to the Spalding-Rigdon theory, participated in writing the *Book of Mormon*. In fact, the Church of Jesus Christ of Latter Day Saints (the Mormon Church) acknowledges that many sections of the *Doctrine and Covenants* were revealed jointly to Smith and Rigdon or to Smith and Cowdery.²⁰ The voice signals of one of these men or a mix of their signals could be the "prophetic voice" Holmes ascribes to Smith. That Holmes would find similarities between the "prophetic" voice of the *Doctrine and Covenants* and the *Book of Mormon*, therefore, is at best evidence of common authorship for the two texts but in no way

demonstrates that Smith's "voice" (divinely inspired or otherwise) is anywhere to be found.²¹

A New Approach

For many, the question of who wrote the *Book of Mormon* remains unresolved. Historical and stylometric research has so far not given us a reliable answer. We offer here a new approach that differs from past work both in source selection and methodology. We examine the entire 1830 *Book of Mormon* without any *a priori* assumptions, modifications or pre-selection, and compare it to new, candidate-author samples. Our methodology does not isolate word categories (i.e. contextual or non-contextual nouns) but instead uses the entire corpus as a starting point and a mathematically based selection process to define the features of the author samples and the *Book of Mormon* that we will compare. Our work employs two techniques to determine the probability that each chapter of the *Book of Mormon* was authored by each of seven authors: Oliver Cowdery, Parley Pratt, Sidney Rigdon, Solomon Spalding, Isaiah-Malachi (from the Bible), Henry Wadsworth Longfellow, and Joel Barlow. The first five have known or alleged connections to the *Book of Mormon*. The last two are prominent, period-authors who were added as controls.²²

The first technique, "delta" (Burrows, 2002, Burrows, 2003, Hoover, 2004a, Hoover, 2004b) is well documented in the literature of computational linguistics, so we omit a detailed description here. The second is "nearest shrunken centroids" (NSC). NSC is a statistical technique for classification in

high-dimensional settings. The problem of authorship attribution is a *classification* problem because we seek to classify a text sample of unknown authorship into one of a fixed number of known author categories - in this case a closed set of candidate authors: Rigdon, Spalding, Cowdery, Longfellow, etc. The problem is *high-dimensional* because we seek to perform classification on the basis of a very large number of words. The method is as follows. First, we compute average word frequency vectors, or *centroids*, for each known author, on the basis of the text samples of known authorship. Next, we shrink these centroids towards the overall average word frequency vector across all of the authors, in order to make our method more robust to small changes in word frequencies. Finally, we classify a text of unknown authorship by computing its word frequency vector and determining to which of the shrunken centroids it is most similar. NSC was initially intended for a completely different purpose: it was developed to assist cancer diagnosis by classifying patient tumor samples into cancer subtypes based on gene expression measurements. However, from a machine learning perspective, the problem of authorship attribution is surprisingly similar to that of cancer diagnosis: rather than classifying tumors by cancer subtype, we classify texts by author, and instead of using gene expression measurements to perform the classification, we use word frequencies.²³ This creates the seven author categories described above. The same kind of analysis is then done on a *Book of Mormon* chapter, and the resulting pattern is compared to the pattern of each of the seven potential author categories (Rigdon–90%; Longfellow–1%; etc.) On this basis, NSC assigns a probability that each potential author wrote

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

each *Book of Mormon* chapter; just as it would assign a probability that a tissue sample manifests a particular cancer sub-type. More details regarding NSC can be found in Tibshirani (2002, 2003).²⁴

Source Selection

Because several theories for the origin of the *Book of Mormon* propose multiple authorship, we cannot investigate it as if it were a single unified text written by a single author, but must instead break it into meaningful samples. Smith reportedly dictated the original document in a series of sessions with scribes.²⁵ These scribes allegedly wrote down everything he said, without punctuation or attention to grammatical form. A key scribe, Oliver Cowdery, is alleged to have provided the initial editing before publication. Subsequent editors altered punctuation, improved grammar, eliminated redundant phrases, and, in some cases, made changes in the text's content. Our investigation thus begins by excluding any analysis of punctuation (e.g. comma frequency) or of form (verse length, sentence length, *etc*), and is limited to words alone. Since it was the published 1830 version of the text that was authorized by the Mormon Church, we examine the entire text, excluding only the chapter summaries that appear before First Nephi, Second Nephi, Jacob, Alma, Helaman, Third Nephi, Forth Nephi, and Ether.

We opted to use the chapter structure currently recognized by modern Mormon Church editors to create our text samples. This results in a total of 239 text segments for testing. This approach yields texts that are generally of

adequate size (verses are too small and books too large), recognizes natural breaks in the narrative, facilitates cross-referencing to on-line resources,²⁶ and avoids the chance that we have imposed our own bias. We consider it important that this method tests the entire corpus approved by Smith in 1830. *Book of Mormon* samples averaged 1117 words and ranged in size from 95 to 3752 words. Our candidate-author samples were equally varied and ranged from a small sample of 114 words to a large sample of 17,797 words with an average sample size of 2172 words.²⁷

For comparative purposes, we acquired digital versions of the Books of Isaiah and Malachi from the *King James Bible* as well as samples of known writings of Solomon Spalding, Sidney Rigdon, Parley Pratt, and Oliver Cowdery (Appendix A provides a detailed list of source materials). For control purposes, we selected two texts: Henry Wadsworth Longfellow's *Song of Hiawatha* (1855) and Joel Barlow's *Columbiad* (1825). Barlow and Longfellow were initially selected as control authors because both are roughly contemporary to the *Book of Mormon*, both deal to an extent with concepts found in the *Book of Mormon*, and both employ formulaic patterns consistent with patterns of verse seen in the *Book of Mormon*.²⁸ To further test the appropriateness of these control texts, we performed a series of simple hierarchical classification tests using frequently occurring non-contextual words and fifty novels of the same era (1789-1850)²⁹. The texts written by Longfellow and Barlow consistently clustered close to the *Book of Mormon* indicating that they were appropriate choices for use as control texts. The Isaiah and Malachi texts also served as *pseudo-control* texts, since

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

large sections of the *Book of Mormon* are known to be almost verbatim extracts from them.³⁰ All of the known author samples were segmented in order to obtain estimates of the variance associated with each author's word use. In total 239 chapters of the *Book of Mormon* and 217 samples of known authorship were tested. Using scripts developed for this project, each sample was tokenized in order to produce word counts and relative frequency data for each word within each sample.³¹ We did not include Joseph Smith in the analysis because, as noted above, there is currently no reliable corpus of Joseph Smith text.

Methodology

As described in the previous section, our data consist of 239 samples of unknown authorship (corresponding to chapters from the *Book of Mormon*) and 217 samples written by seven known authors. We refer to this analysis as the “seven-author case.” The number of text samples used for this analysis are as follows: Cowdery (nineteen), Pratt (fifty-three), Rigdon (twenty-three), Spalding (seventeen), Isaiah-Malachi (seventy), Barlow (twelve), Longfellow (twenty-three). We used a set of 110 words or “features,” obtained in three steps:

- 1) We selected the words that occurred at least once in the samples from each author and also at least once in the *Book of Mormon*. This resulted in a set of 521 words.
- 2) We selected the subset of these 521 words that have a mean relative frequency, across the 456 samples, of at least 0.1%. This resulted in a set of 114 words.
- 3) We removed the words “god”, “ye”, “thy”, and “behold”, as these occurred at much higher frequencies in texts relating to biblical subject matter.

The resulting list of 110 words is available in Appendix B.³²

In order to compute delta scores and apply NSC, we first converted the 110 word counts for each text into relative word frequencies. For NSC, we formatted the data as a matrix of dimension 456x110 (number of samples by number of words). We subtracted out the mean from each column and divided the entries in each column by the standard deviation for that column. We then applied NSC to the data, using the “pamr” (Prediction Analysis for Microarrays) package that is freely available on the R statistical software website.³³

Both delta and NSC involve the selection of tuning parameters. For both methods, this tuning parameter determines the number of words to include in the classifier. In order to determine the success rates of NSC and delta at classifying chapters of known authorship, and in order to select a value for the tuning parameters, we performed cross-validation. Roughly speaking, cross-validation is performed as follows, for a range of values of the tuning parameter:

- 1) Randomly split the samples of known authorship into two sets: a “training set,” containing most of the samples, and a “test set,” containing a smaller portion of the samples.
- 2) Perform the classification method of interest (either delta or NSC) for a given value of the tuning parameter, training on the training set and testing on the test set.
- 3) Compute the error fraction from the number of misclassified test set samples.

Cross-validation allowed us to estimate the error that we would obtain if we tried to classify the samples of *known* authorship using NSC and delta. The above process was repeated multiple times, and the average misclassification error rate recorded. The lowest delta error rate of 11.1% was obtained using ninety words.

This means that if we used delta to classify a new sample written by one of the

seven known authors, then the probability of correct classification would be 88.9%. The lowest NSC error rate was obtained when all 110 words were included; the error rate was 8.8%. This means that we would expect to classify correctly a new sample written by one of the seven candidate authors 91.2% of the time. Since there are seven candidate authors, a classifier that selected an author completely at random would give a correct classification rate of $1/7$ or 14.3%, and an average misclassification error rate of $6/7$, or 85.7%. Therefore, the low error rates obtained using NSC and delta are impressive. The fact that NSC results in lower error rates indicates that this method is appropriate for authorship attribution, and may in this case be superior to delta.

Using delta, five of the 239 chapters of the *Book of Mormon* were incorrectly assigned to control author Longfellow (none to Barlow), an error rate of 2.1%. Using NSC, only two chapters were assigned incorrectly to Longfellow (none to Barlow), an error rate of 0.8%. To provide best estimates of individual chapter authorship for the five authors who are linked historically to the *Book of Mormon* (Spalding, Rigdon, Cowdery, Pratt) or who are known to have contributed (Isaiah-Malachi), we also performed a second delta and NSC analysis (hereafter referred to as the “five-author case”) in which we omitted the Barlow and Longfellow control texts. In the five-author case, the lowest NSC error rate was obtained using 108 words (listed in Appendix B).

Results

For each chapter of the *Book of Mormon*, using both NSC and delta, we compared the relative probability that a candidate author or a control author contributed to that chapter. We then established a "ranking" for each of the seven authors (1-7) from most likely to least likely and calculated the percentage point difference between candidates in terms of their probability. In Alma forty-seven (chapter 147), for example, the first place ranked candidate (using NSC) has a probability of 46.5% where the second place candidate is 46.3%. Given this close proximity it would be impossible to conclude that one candidate is more likely than the other. In the majority of chapters, however, we do not observe this sort of close probability between first and second candidates. Most chapters (57%) show at least a fifty-percentage point difference between first and second choice. Indeed, in forty chapters (17%), the difference between first and second most probable author is over ninety percentage points. Second Nephi twenty-two (chapter forty-four), for example, is a chapter known to contain strong borrowings from the Book of Isaiah. NSC ranks the probability of Isaiah-Malachi as the source for this chapter at 99.99%. In fact, twenty of the twenty-one chapters known to have been borrowed from Isaiah or Malachi are properly attributed at a probability at or above 91% certainty.³⁴ There was thus only one "false negative" for chapters that are known to be derived from Isaiah-Malachi (Mosiah fourteen is borrowed from Isaiah fifty-three but was attributed to Longfellow). This is evidence for the effectiveness of NSC classification. Further evidence comes from a consideration of "false positives"—chapters attributed

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

incorrectly to Isaiah-Malachi. There are twenty-one known Isaiah chapters and another sixteen that have some relationship to Isaiah or Malachi (about 15% of the chapters in the *Book of Mormon*). But Delta assigns 47% to Isaiah-Malachi, while NSC assigns 27%. This indicates that both delta and NSC had “false positives” for Isaiah-Malachi, but the NSC false positive error rate was about half that of delta.

NSC and Delta agree on the first place assignment for 147 of 239 chapters (62% agreement). In cases where there is not first place agreement between the two methods, there are seventy-six chapters in which the first place candidate of one method agrees with the second place candidate in the other. There are a total of 223 chapters (93%) in which the two methods name the same author in either the first or second place. In the 147 chapters where both methods agree on first place, there are two chapters assigned to Cowdery, two to Longfellow, four to Pratt, thirty-four to Spalding, forty-six to Rigdon, and fifty-nine to Isaiah-Malachi. In the seventy-six chapters where there is agreement between a first choice in one method and a second choice in another, there are forty-two cases, which are inverses of each other, that is there are forty-two cases in which the author listed as first place in one method is listed in second place in the other method. In these instances, there are nine cases in which Rigdon is paired with Cowdery, twenty cases in which Rigdon is paired with Isaiah-Malachi, eight cases in which Rigdon is paired with Spalding, two cases in which Spalding is paired with Pratt, and three cases in which Spalding is paired with Isaiah-Malachi.

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

Examining the NSC results (see Table 1), we note the following for the most likely positions of the first and second most probable candidate: Rigdon is either the first or second most probable candidate author in 197 out of 239 chapters; Spalding is either the first or second most likely candidate in 110 out of 239 chapters; and Isaiah-Malachi is first or second in 101 chapters. Cowdery appears thirty-seven times in first or second place and Pratt appears twenty-four times. Barlow is never seen in first place and appears only once in second place. Longfellow is first in just two chapters and second in only six. Additionally, we note that for the least likely positions of the sixth and seventh most probable candidates, Rigdon never shows up in last (seventh) place and appears only twice in sixth place. Spalding never shows up in either seventh (last) or sixth place. Isaiah-Malachi appears four times in last place and fourteen times in sixth. Cowdery appears 116 times in sixth or seventh place, Pratt thirty-four times. Barlow appears in the sixth or seventh place 175 times and Longfellow 133 times. Table 2 shows similar results generated from the delta classification.

Table 1. Number of chapters assigned to each author based on NSC probability assignments for the seven-author case (Barlow and Longfellow controls included). If chapter assignments were random, the expected number for each choice would be $239/7 = 34$.

Rank	Cowdery	Pratt	Rigdon	Spalding	Isaiah & Malachi	Barlow	Longfellow
1 st choice	20	9	93	52	63	0	2
2 nd choice	17	15	104	58	38	1	6
3 rd choice	26	49	29	59	56	5	15
4 th choice	31	51	7	64	39	18	29
5 th choice	29	81	4	6	25	40	54
6 th choice	16	27	2	0	14	99	81
7 th choice	100	7	0	0	4	76	52

Table 2. Number of chapters assigned to each author based on delta probability assignments for the seven-author case (Barlow and Longfellow controls included). If chapter assignments were random, the expected number for each choice would be $239/7 = 34$.

Rank	Cowdery	Pratt	Rigdon	Spalding	Isaiah & Malachi	Barlow	Longfellow
1 st choice	5	7	63	47	112	0	5
2 nd choice	19	22	92	23	54	9	20
3 rd choice	53	26	31	50	39	15	25
4 th choice	23	50	24	70	16	22	34
5 th choice	34	59	18	39	11	27	51
6 th choice	32	63	10	10	5	48	71
7 th choice	73	12	1	0	2	118	33

Summing the probabilities for each candidate-author across the entire *Book of Mormon*, allows visualization of the relative presence of each candidate-author's "signal" in the overall text (see figures 1 and 2 below).

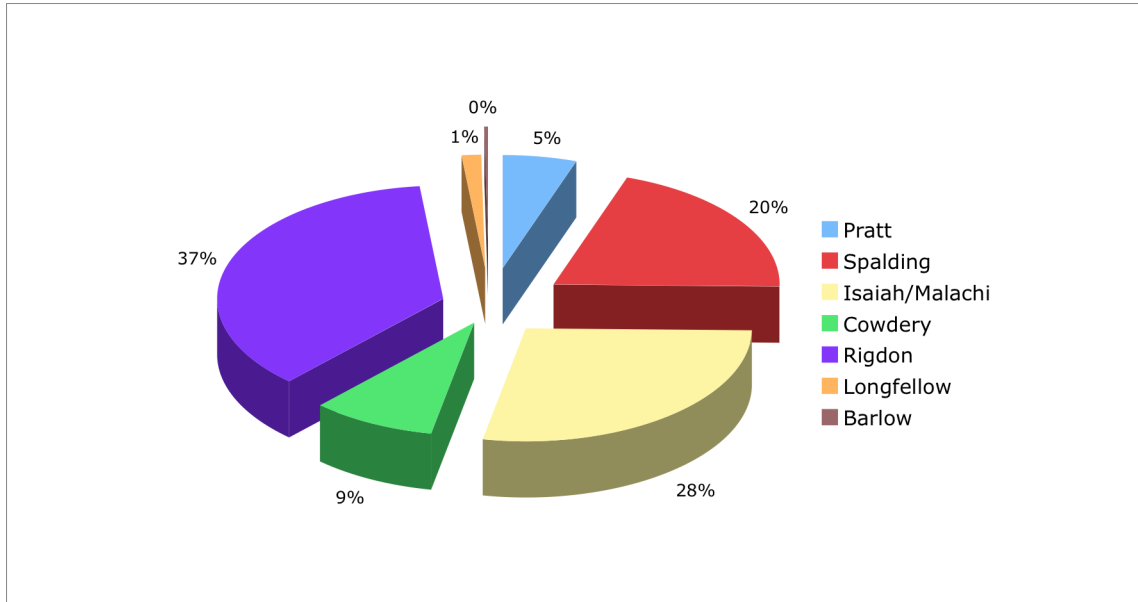


Fig. 1 Overall attribution percentages as assigned by NSC

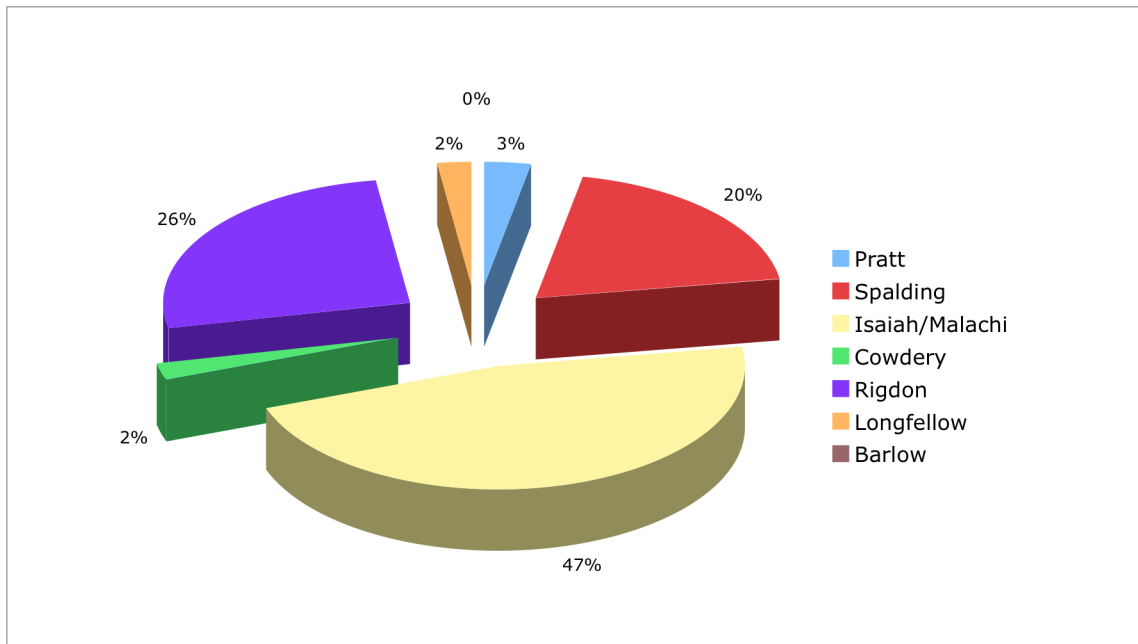


Fig. 2 Overall attribution percentages as assigned by delta

In both classification methods, the signals for Rigdon, Isaiah-Malachi, and Spalding are dominant and the signals for control authors Longfellow and Barlow

are comparatively small or altogether absent. The Pratt and Cowdery signals are present but small beside the signals for Rigdon, Isaiah-Malachi and Spalding. Both NSC and delta tend to agree closely in terms of the relative presence of the Rigdon and Spalding signals. The greatest disagreement between the two methods appears in relation to the Isaiah-Malachi signal where delta assigns 47% of the chapters to Isaiah-Malachi as a first choice while NSC assigns 28%. The actual Isaiah-Malachi percentage can be estimated at around 36 chapters, or 15% of all chapters. In other words, 15% of the *Book of Mormon* is derived from Isaiah-Malachi or contains excerpts from Isaiah-Malachi. This indicates that while both delta and NSC had false positives, NSC had many fewer and is closer to the actual or true value.

Figures 3 and 4 show the number of chapters assigned to each author as either the first or second most likely attribution. Again, we note the dominance of Rigdon, Isaiah-Malachi, and Spalding in both first and second place assignment and the comparatively small presence of both the control authors and the other candidates.

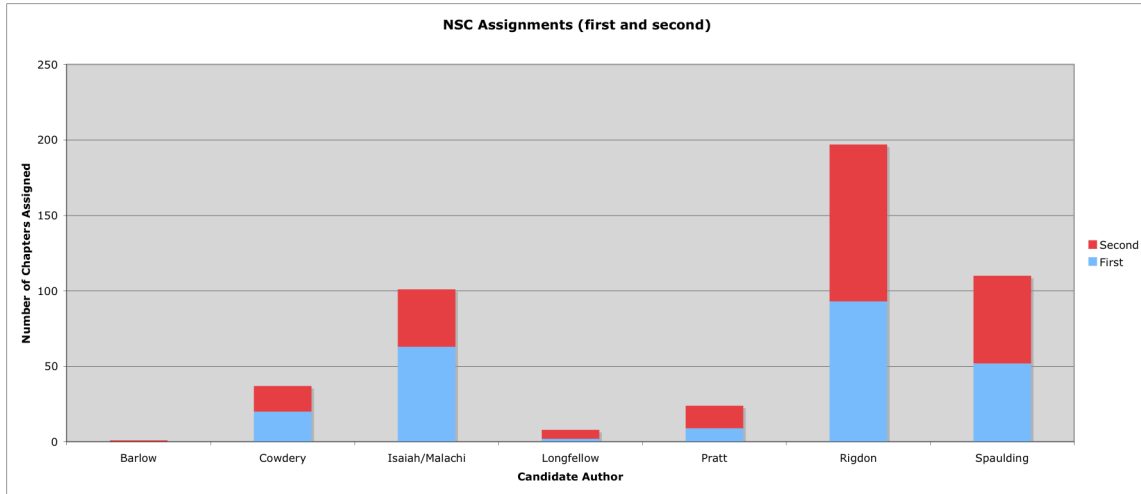


Fig. 3 First and second place NSC assignments for the seven-author case

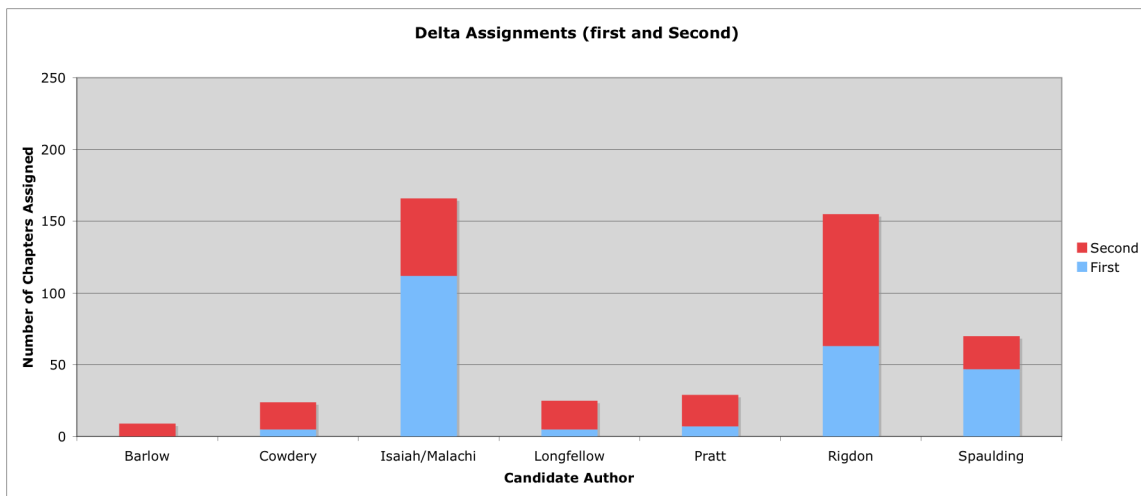


Fig. 4 First and second place delta assignments for the seven-author case

All of the above results are for the seven-author case. The five-author case gave highly similar results. For the first most likely attribution, identical results were obtained for 226 of the 236 chapters (96% agreement). For the first and second most likely attributions, identical results were obtained for 223 of the 239 chapters (93% agreement).

Discussion

In the cross-validation tests, NSC was more effective, so our discussion and conclusions are based on the NSC data unless otherwise specified. Use of NSC enabled assignment of author probabilities based on the closeness of each chapter within the *Book of Mormon* to the known linguistic signals of a set of candidate nineteenth-century authors. These probabilities in turn made it possible to gauge the relative presence of one signal over another. The low signals for the control texts indicate that NSC effectively identified and did not select the control authors (see figures 5 and 6).

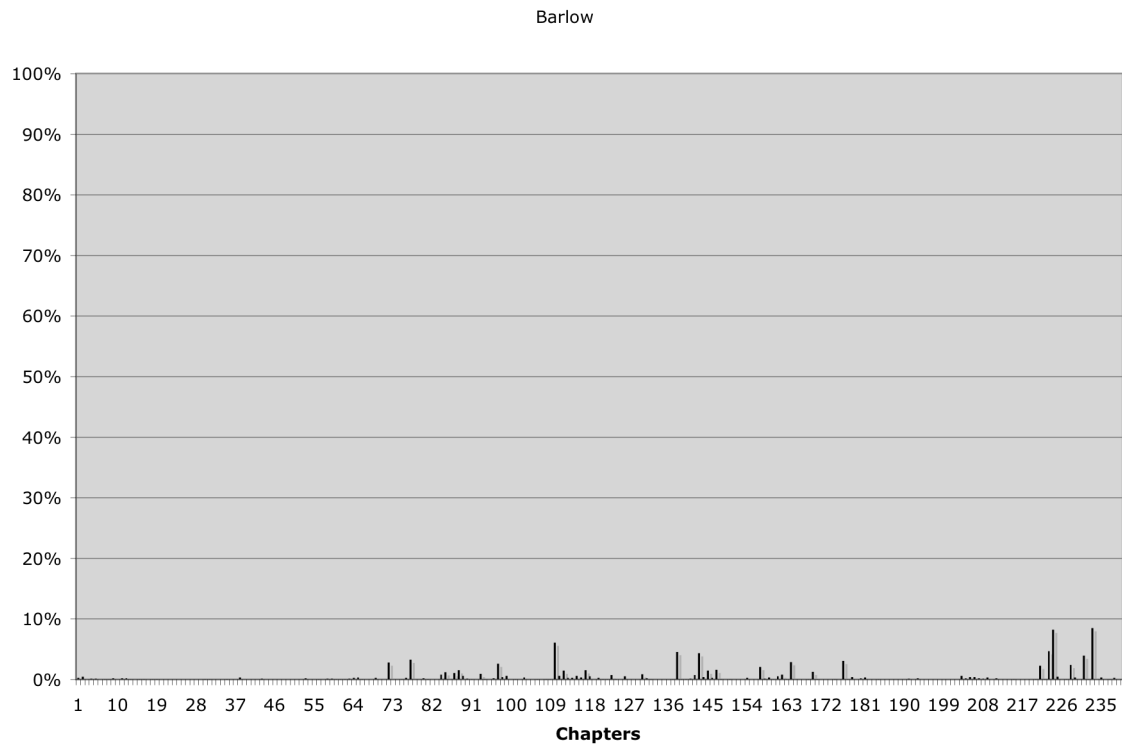


Fig. 5 Chapter-by-chapter probability of Joel Barlow as author

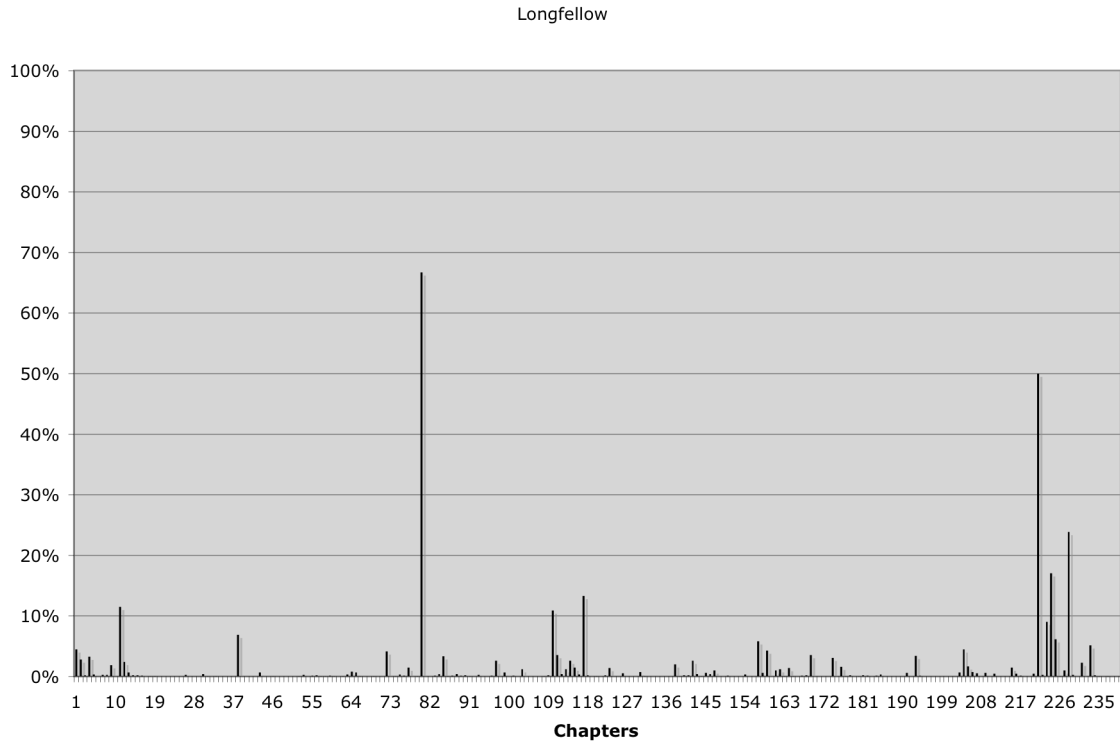


Fig. 6 Chapter-by-chapter probability of Henry Wadsworth Longfellow as author

At a macro level, the signals for Rigdon, Isaiah-Malachi, and Spalding dominate the *Book of Mormon*, with NSC assigning 85% of the text to one of these three candidate authors (delta assigns 93% to these three). Taken together, Rigdon and Spalding account for 57% percent of the first place assignments and 68% of the second place assignments. Isaiah-Malachi accounts for 28% of the first place assignments and 16% of the second place assignments. Overall only ten chapters lack a signal for Rigdon or Spalding in one of the two most probable positions and four of these ten chapters are chapters known to be derived from the Book of Isaiah, as discussed below. In other words, of the 239 chapters in the *Book of Mormon*, 229 show either the

Rigdon or Spalding signal prominently. Together Rigdon and Spalding receive 64% of the combined first and second place assignments, Isaiah-Malachi receives 21% and other candidates or control-authors receive 15% (figure 3).

It is well-accepted that some chapters from the Old Testament books of Isaiah and Malachi served as source material for the *Book of Mormon*. Both delta and NSC correctly classified all twenty-one chapters of the *Book of Mormon* that contain strong borrowings from Isaiah and/or Malachi, and all were classified with a probability at or above 84% by NSC.³⁵ That said, both methods also detected the Isaiah-Malachi signal in chapters that are not obviously derived from the Old Testament. For NSC, there are forty-three such chapters and, in thirty-six of these, the probabilities strongly favor Isaiah-Malachi over the other possible candidates.³⁶ These are chapters where the style and word usage patterns are close to those in the Old Testament. Others (Palmer, 2002, Pursuite, 2000, Tanner, 1998, Walters, 1990, Marquardt, 2000) have spent considerable time tracing the direct correspondences between the *Book of Mormon* and the King James version of the Bible, so we will not delve into the specifics of these Isaiah-Malachi attributions other than to note that fifteen of these thirty-six chapters are directly related to Isaiah and one of the thirty-six is a chapter borrowed from the New Testament.³⁷ Discounting the sixteen chapters that have some connection to the Bible, this leaves twenty chapters attributed to Isaiah-Malachi for reasons that are not obvious. We note, however, that many of these twenty chapters have thematic similarity to Ethan Smith's *View of the Hebrews* (Pursuite, 2000, Walters, 1990), which is believed by many to be linked to the *Book of Mormon*

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

through Cowdery,³⁸ and a future analysis might utilize Ethan Smith's text as a potential source. Figure 7A shows the relative presence of the Isaiah-Malachi signal across the entire *Book of Mormon* and figure 7B shows chapters attributed to Isaiah-Malachi.

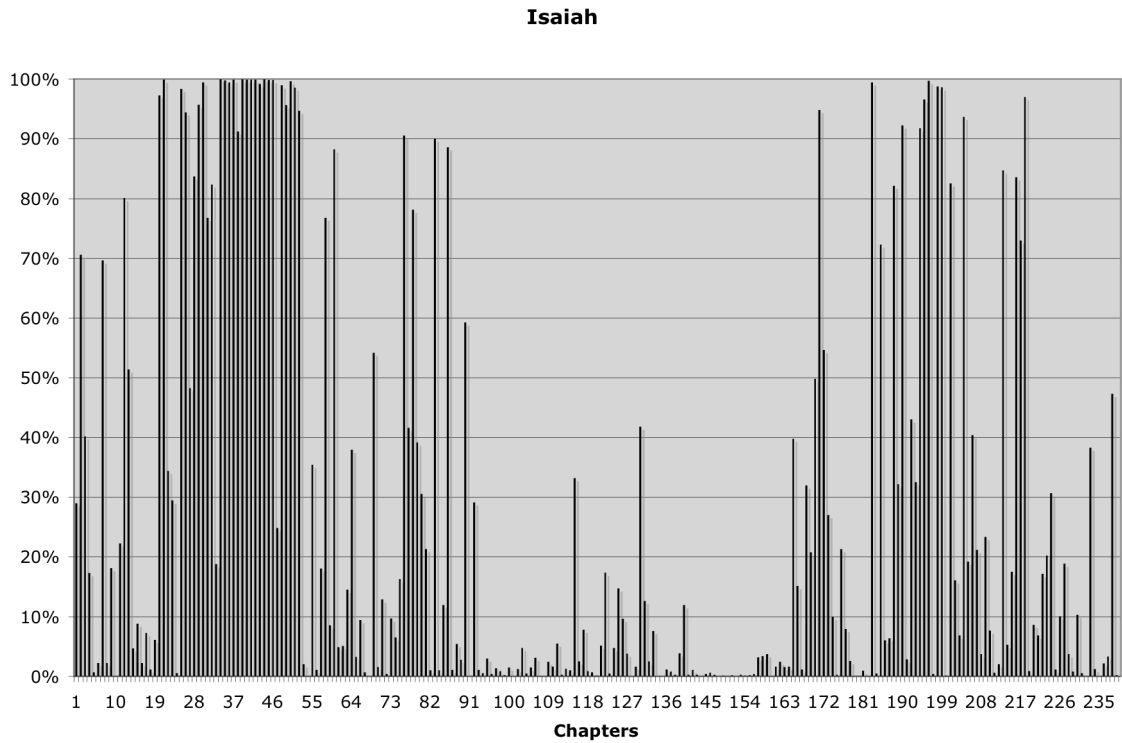


Fig. 7A Chapter-by-chapter probability of Isaiah-Malachi as author (seven-author case)

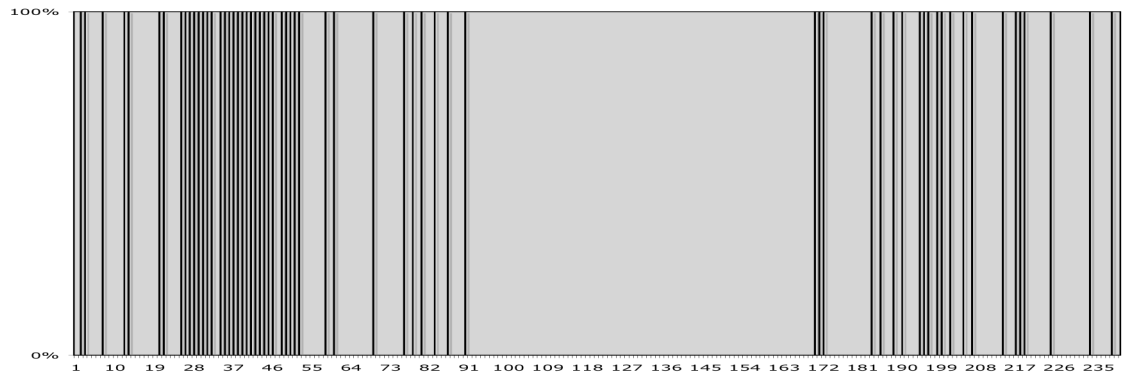


Fig. 7B Chapters for which the first place NSC assignment was Isaiah-Malachi (five-author case)
 DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

The prominence of the Rigdon and Spalding signals are significant and provide strong support for the Spalding-Rigdon authorship theory: that Rigdon acquired one or more manuscripts written by Spalding and then modified them, by incorporating his own theology, to create the 1830 version of the *Book of Mormon*. Figure 8A illustrates the presence of the Rigdon signal through each chapter of the *Book of Mormon*, and Figure 8B shows chapters attributed to Rigdon. The graph shows a dominant Rigdon signal in First Nephi, the non-Isaiah fraction of Second Nephi, Jacob, Enos, Words of Mormon, Mosiah, Helaman, the non-Isaiah fraction of Third Nephi, Mormon, Ether, and Moroni, with an intermittently strong signal in the Book of Alma. Especially noteworthy here is the fairly regular distribution of the signal across the entire text. A gap in the Rigdon signal appears in sections known to be copied from Isaiah (Figure 7), in portions of the book of Alma attributed to Spalding (Figure 9), and it is sporadic in the first quarter of the text, in the section known to scholars of Mormonism as replacement material added after Smith's loss of 116 pages he claimed to have translated.³⁹ The lost pages contained material that would have ended near the beginning of the Book of Mosiah. It is generally held that Smith resumed his purported translation at Mosiah and continued through to the end of the *Book of Mormon*, returning at the end of the process to replace the lost pages. One possible scenario is that Smith and/or Rigdon prepared a replacement in fall of 1828 by drawing from source material at hand such as the Book of Isaiah (which

features prominently in this part of the text) and perhaps from Ethan Smith's *View of the Hebrews*.

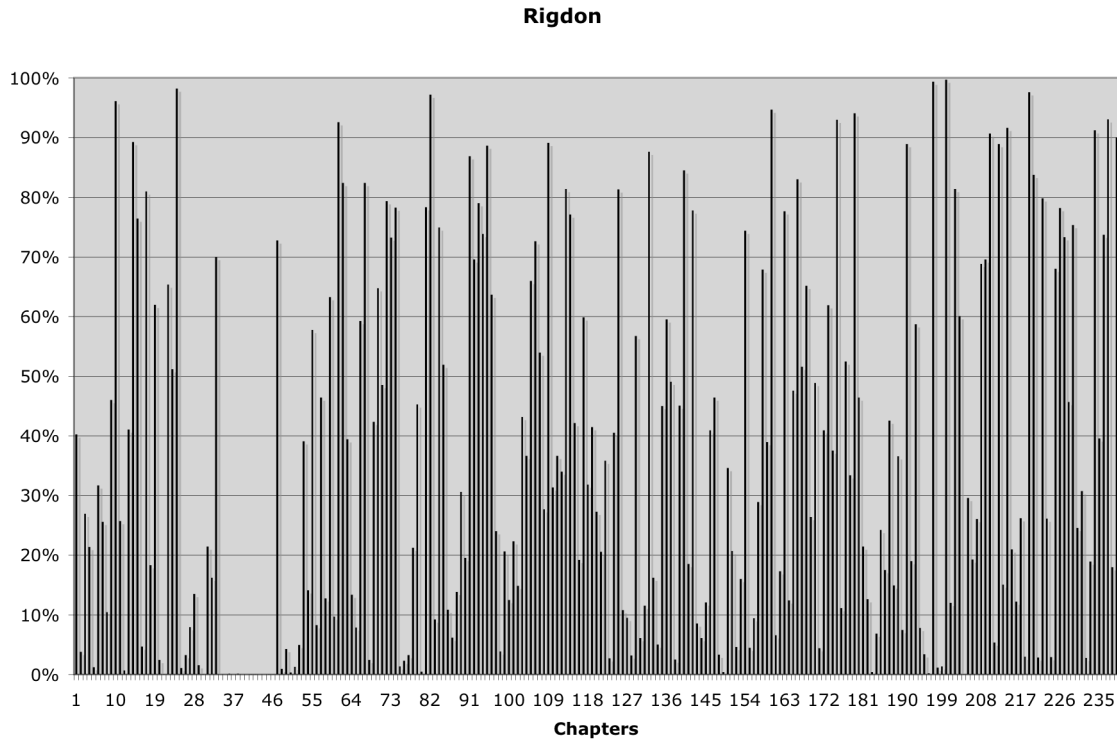


Fig. 8A Chapter-by-chapter probability of Sidney Rigdon as author (seven-author case)

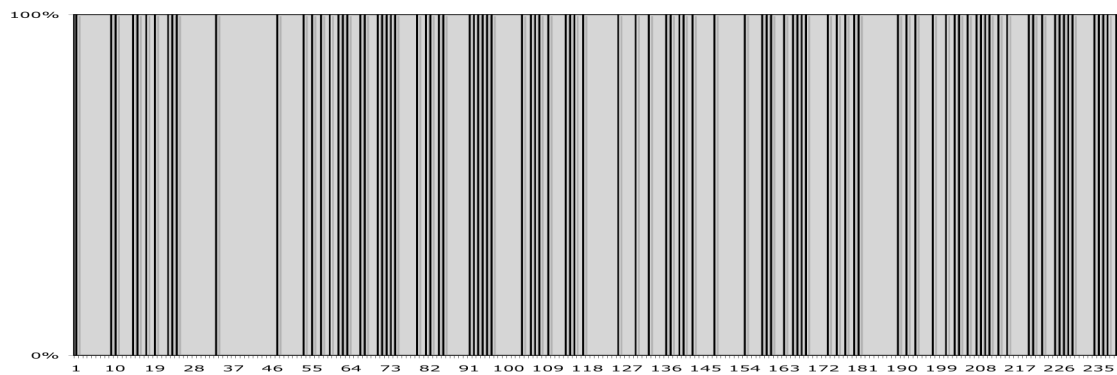


Fig. 8B Chapters for which the first place NSC assignment was Rigdon (five-author case)

Figure 9A illustrates the presence of the Spalding signal through each chapter of the *Book of Mormon* and 9B the chapters attributed to Spalding. Noteworthy here is (1) the small Spalding signal in sections of the *Book of Mormon* that were likely added to replace the 116 pages (i.e. the first quarter of the book - First Nephi through Words of Mormon), and (2) the fact that the chapters with a dominant Spalding signal are primarily narrative and non-theological, and thus consistent with descriptions of "Manuscript Found," the missing Spalding document that is alleged to be foundational to the *Book of Mormon* (Howe, 1834, 1977). The prominence of the Spalding signal in the Book of Alma is especially noteworthy. Dale Broadhurst has identified these chapters as likely Spalding contributions based on his careful comparison of phrases found both in the Book of Alma and Spalding's "Manuscript Story."⁴⁰ Similar thematic and linguistic patterns between the *Book of Mormon* and Spalding's "Manuscript Story" have also been identified by Holley (1989).⁴¹

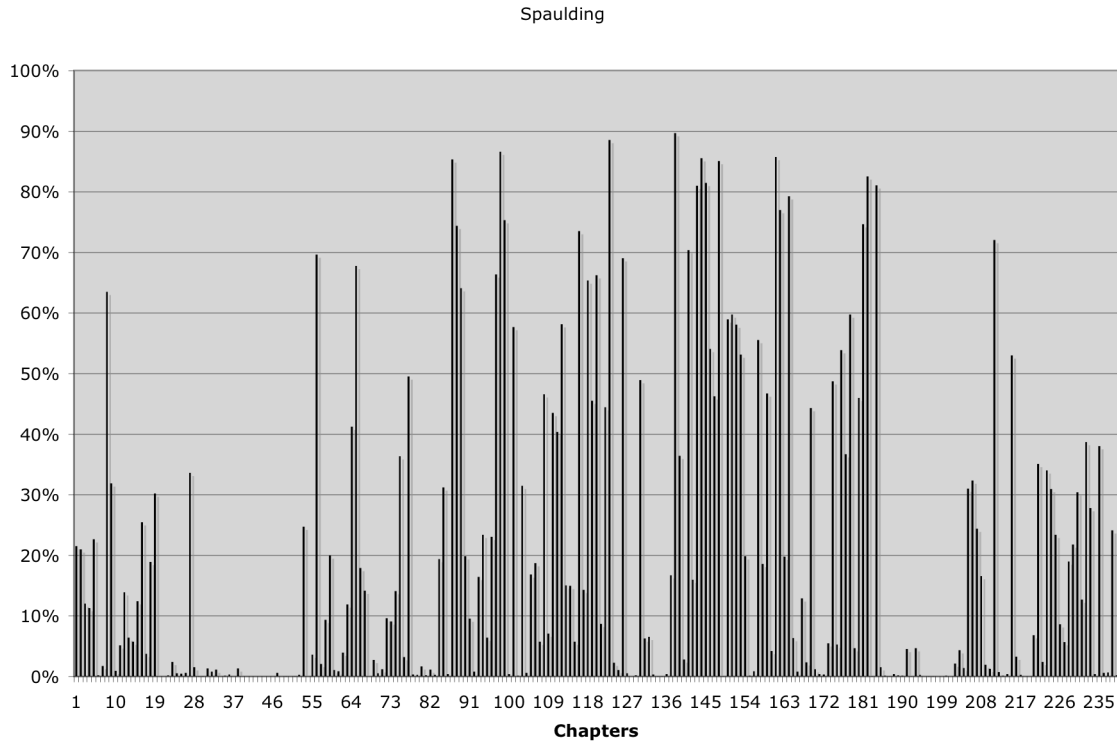


Fig. 9A Chapter-by-chapter probability of Solomon Spaulding as author (seven-author case)

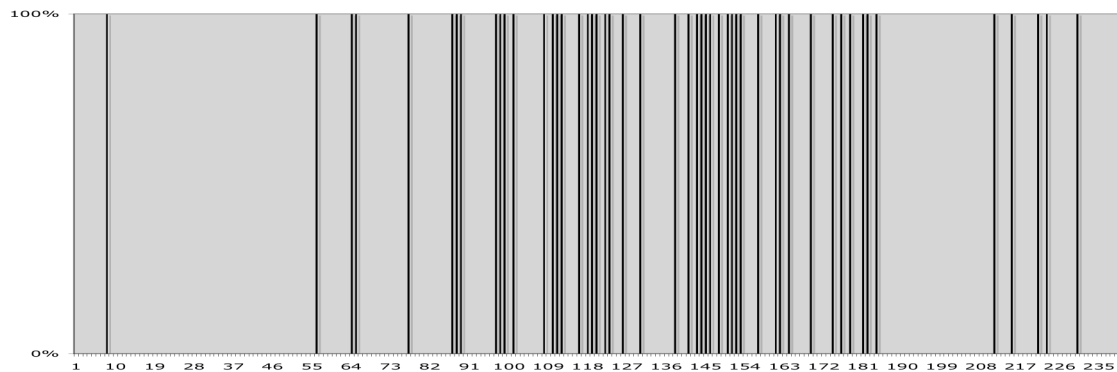


Fig. 9B Chapters for which the first place NSC assignment was Spaulding (five-author case)

Figure 10A shows the distribution of the Oliver Cowdery signal and 10B the chapters attributed to Cowdery.⁴² The Cowdery signal is most prominent in the middle third of the book with a strong cluster of authorial assignments

(fourteen first-place and four second-place) in the Book of Alma. Where
 DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using
 Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic
 Computing*, December 2008; 23: 465 - 491

Cowdery is the most probable author, he is paired with Rigdon as second most probable author in all but two cases; where Cowdery is assigned as second most probable author (seventeen chapters), Rigdon is first most likely in fourteen of these. All of this suggests a strong correlation between Cowdery and Rigdon and the likelihood that if Cowdery contributed to the *Book of Mormon*, he may have done so in collaboration with Rigdon. The Cowdery signal appears only where the Rigdon signal is also prominent and in many cases the difference between the strength of the two signals is marginal. Also noteworthy is that the Cowdery signal appears most prominently in the middle third of the book. His signal appears after the Book of Mosiah and near the beginning of the Book of Alma—the point in the manuscript where Smith supposedly began to dictate with Cowdery as his scribe, and when the speed of translation reportedly increased significantly.⁴³ It is in these sections of the *Book of Mormon*, especially the third quarter of the Book of Alma, that we find the Cowdery signal—in well-composed chapters that deal with such topics as the nature of faith (Alma thirty-two), atonement through Christ⁴⁴ (Alma thirty-six), and liberty (Alma sixty-one). Still, if Cowdery had a direct hand in the authorship of the *Book of Mormon* it was likely a lesser one.⁴⁵ It is more likely that his primary role was editorial given both the historical and stylometric data.

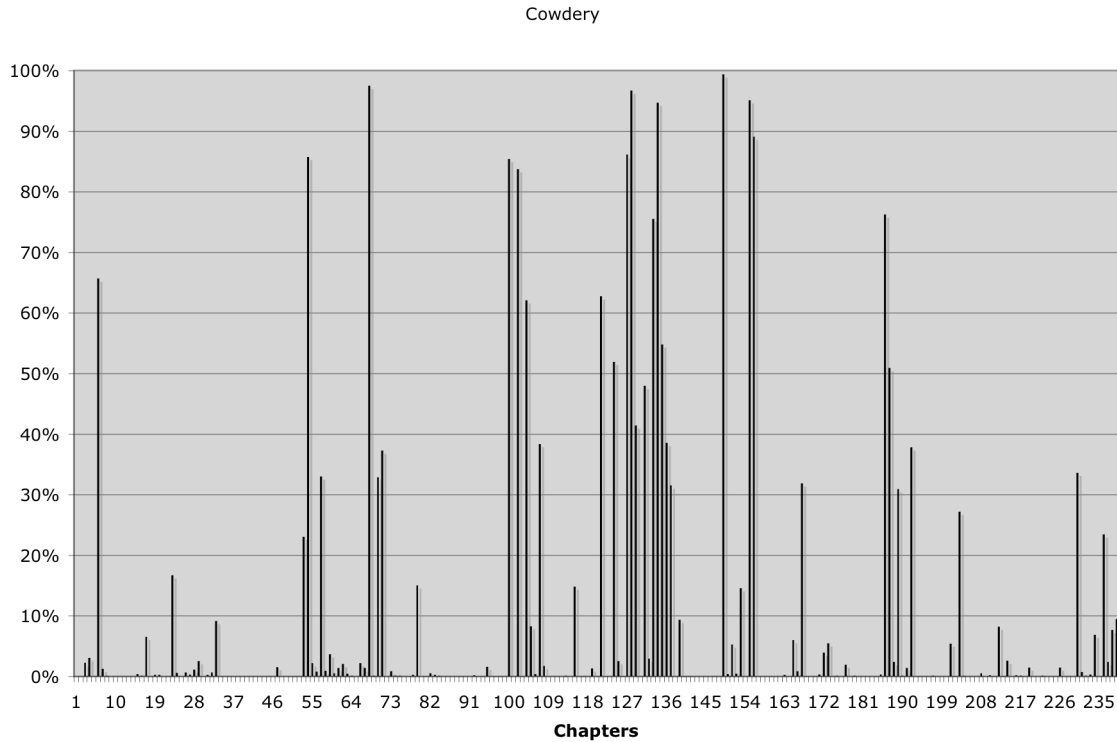


Fig. 10A Chapter-by-chapter probability of Oliver Cowdery as author (seven-author case)

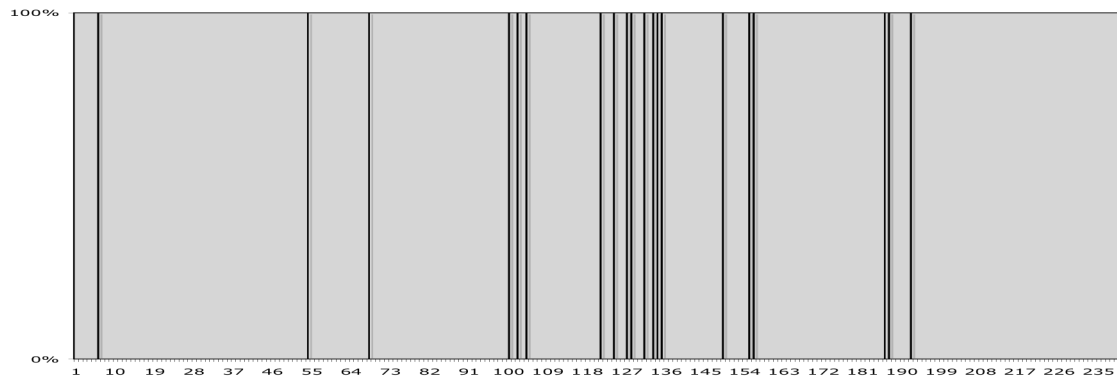


Fig. 10B Chapters for which the first place NSC assignment was Cowdery (five-author case)

Figure 11A shows the distribution of the Parley P. Pratt signal and 11B the chapters attributed to Pratt.⁴⁶ Pratt is the most likely author for nine chapters with five occurring in First Nephi, one in Mosiah, and two small chapters appearing, back-to-back, in Moroni (Figure 11, Panel B). Pratt was an early leader in the

Mormon church and one of the original Quorum of the Twelve Apostles. In 1826, however, he was a wandering tin peddler who "knew everybody in Western New York and Northern Ohio" (Schroeder, 1901, Shook, 1914). He lived near Rigdon's residence in Bainbridge, Ohio, and joined Rigdon's congregation.⁴⁷ During the same period Rigdon is reported to have collaborated with "two or three different persons" in "adjacent places" to create the *Book of Mormon*.⁴⁸ Sometime around 1827, Pratt decided to sell all his goods and take up the ministry. It has been suggested that Pratt was "the medium through whom Rigdon made the acquaintance of Smith when seeking a suitable tool for his purpose" (Eaton, 1882, Williams, 1842).⁴⁹ While traveling in 1830, ostensibly to see family, Pratt reported sudden inspiration that led him to Palmyra, New York, where he quickly converted to Mormonism and was baptized by Oliver Cowdery. He and Cowdery then reportedly delivered a copy of the published 1830 version of the *Book of Mormon* to Rigdon. Pratt's conversion is described in contradictory accounts, as is his role in delivering the *Book of Mormon* to Rigdon (Schroeder, 1901).

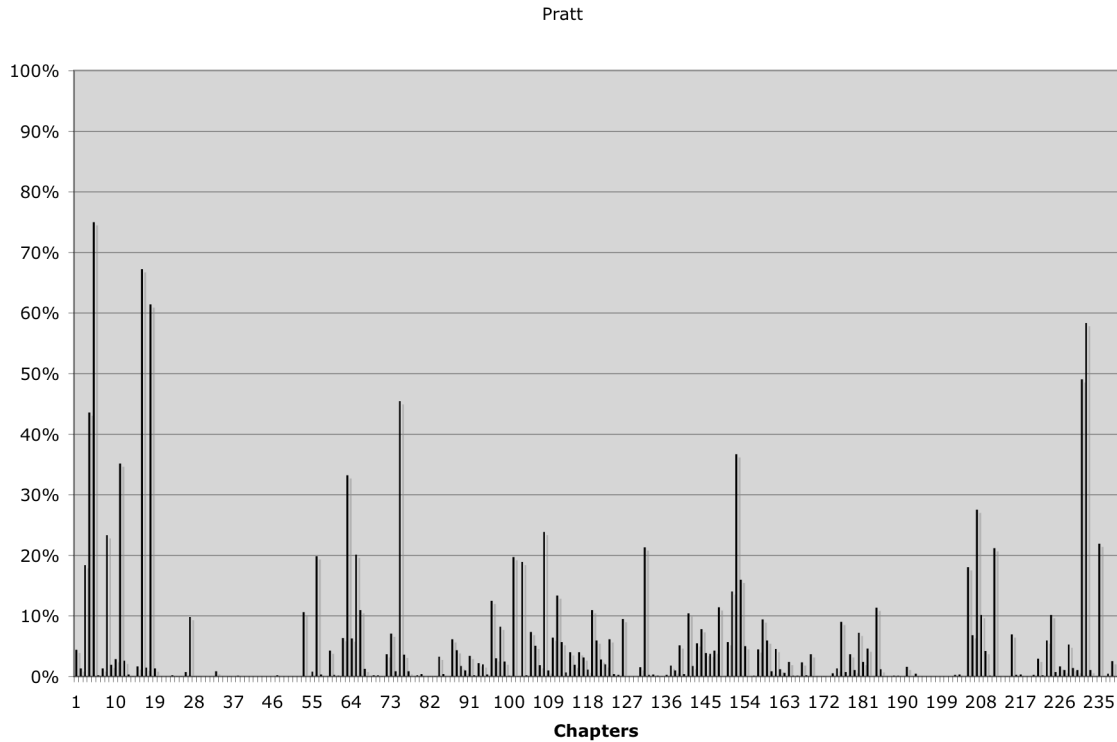


Fig. 11A Chapter-by-chapter probability of Parley Pratt as author (7 author case)

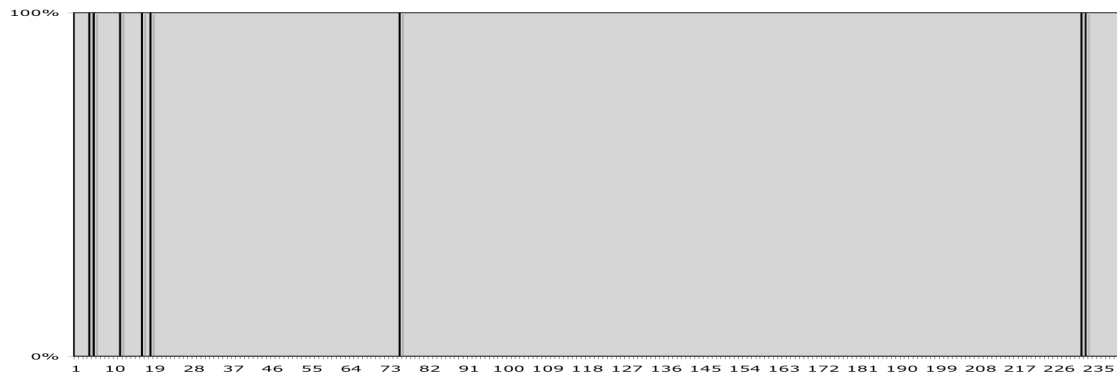


Fig. 11B Chapters for which the first place NSC assignment was Pratt (five-author case).

In five of the nine chapters attributed to Pratt, Pratt is paired with Spalding in second place and in four with Rigdon in second place. Pratt receives fifteen second-place assignments: most of them (ten) as a second to Spalding, and three as a second to Rigdon. The largest proportion (one-third) of the

assignments to Pratt as the second most probable author occurs in Alma, and there are two cases in First Nephi. If Pratt contributed to the *Book of Mormon*, he played a minor role and was likely most involved in First Nephi, where there are several first and second place Pratt assignments.

In the stylometric studies cited earlier, Larson et al. (1980) and Hilton (1988) attempted to test the hypothesis that the *Book of Mormon's* purported ancient authors had dissimilar writing styles. Recent studies in cultural and linguistic evolution suggest another relevant hypothesis by demonstrating that writing styles in ancient texts tend to become increasingly divergent over time (Farmer, 2006). Our chapter-by-chapter analysis tested both hypotheses and found that the *Book of Mormon* does not display patterns consistent with the type of ancient record it purports to be. For example, two of the *Book of Mormon's* alleged principal authors were Nephi and Moroni. They allegedly lived about 1000 years apart. NSC assigned many of their chapters to Rigdon. For example, NSC assigned both First Nephi ten and Moroni eight to Rigdon with >93% probability. The *Book of Mormon* also attributes many chapters to a single ancient author, but our results frequently disconfirmed this. For example, where the *Book of Mormon* attributes Mormon five, six, and seven to an ancient author named Mormon, NSC assigned chapters five and seven to Rigdon (89 and 92% probability, respectively) and chapter six to Spalding (72% probability). Chapters five and seven contain references to the future redemption of the House of Israel, a concept popular in the early nineteenth century and embraced by Rigdon, while chapter six is a war narrative similar to other such narratives penned by

Spalding, a veteran of the American Revolutionary War. These results stand in contrast to claims that the *Book of Mormon* is of ancient authorship.

Conclusions

NSC has proved highly useful for authorship classification. It has a lower cross-validation error rate than delta, a lower rate of false positive assignments, and a probability-based output that enabled in-depth interpretation of the results, including speculation regarding possible connections between candidate authors. The NSC results are consistent with the Spalding-Rigdon theory of authorship. Evidence supporting this conclusion includes the prominence of signals for Spalding and Rigdon; the presence of strong Spalding signals in sections of the *Book of Mormon* previously linked to Spalding; the presence of a dominant Rigdon signal in most theological sections, and a strong Spalding signal in the more secular, narrative sections. Our findings are consistent with historical scholarship indicating a central role for Rigdon in securing and modifying a now-missing Spalding manuscript. The high number of Spalding-Rigdon pairings in first and second place strongly suggests that Spalding and Rigdon were responsible for a large part of the text. Pearson's chi-square test of independence was performed and indicates that the distribution of first-place assignments is significantly different from uniform ($p < 2 \times 10^{-16}$). Similarly, the distribution of second-place assignments differs significantly from uniform ($p < 2 \times 10^{-16}$). Clearly far more chapters are attributed to Rigdon, Spalding and Isaiah-Malachi than might be expected due to mere chance. Other connections

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

detected through this work are also consistent with the historical record, including the likelihood of a lesser, largely editorial role for Cowdery and a possibly minor, if unexpected, role for Pratt.

Based on this evidence we find the original claims of Howe (Howe, 1834, 1977) and the more recent assertions of Cowdrey and coworkers quite plausible; it seems likely that the 1830 version of the *Book of Mormon* was the creation of Sidney Rigdon, a Reformed Baptist Preacher, who had motives, means, and opportunity to carry out the project (Cowdrey et al., 2005). We acknowledge that because our samples of Rigdon prose all come after 1830, some could argue that Rigdon's prose was influenced by the *Book of Mormon* and not *vice versa*. To raise such an objection, however, one would have to argue that Rigdon was so influenced by the *Book of Mormon* that he consciously or unconsciously adopted, even internalized, the most subtle and unremarkable linguistic patterns found in certain portions of the text, but not in others.

Prior exposure to the *Book of Mormon* most certainly did not influence Solomon Spalding who died fourteen years before it was published. Yet our data strongly support the historical claim that a lost Spalding manuscript served as a source text for the backbone narrative of the *Book of Mormon*. The document that we used for samples of Spalding's writing ("Manuscript Story" also known as "The Oberlin Manuscript") does not match the eyewitness descriptions of "Manuscript Found," the draft novel that Spalding read to friends and family in Conneaut, nor does it match the *Book of Mormon*.⁵⁰ The Spalding-Rigdon theory rests heavily on the assumption that additional Spalding manuscripts once

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

existed, and that material from one of these manuscripts provided the narrative framework for the *Book of Mormon*. This additional manuscript would be the one that the Conneaut witnesses and others identified as being the “source” of the *Book of Mormon*. While not that manuscript, the Oberlin Manuscript nevertheless provides us with a reliable sample of Spalding's prose and the linguistic signal detected in it appears with significant regularity throughout the *Book of Mormon*.

Of course, we have not considered every possible candidate-author who may have influenced the composition of the *Book of Mormon*. We have, however selected from among the most likely candidates, excepting perhaps Joseph Smith. In the case of Joseph Smith, we had no reliable samples of prose to test. When reliably identified materials become available, their addition to this analysis would be worth considering. An effort to compile such writings is currently underway⁵¹.

Knowledge of who likely constructed the *Book of Mormon* has significant implications for scholarship in Mormon history and for religious and cultural studies generally, as it addresses the foundation of an emerging world religion now estimated at thirteen million members. 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.⁵²

Acknowledgments

The authors thank David L. Marshall for a transcript of the Rigdon revelations in the Post Collection and for a careful review of a draft of this manuscript; Robert McCue and Denise Gigante for their thoughtful reviews of the manuscript, and Amy Kapp for her early contributions to our thinking about this project.

Notes

¹ Corresponding Author.

² For Smith, see <http://www.lds-mormon.com/campbell.shtml> for Rigdon
<http://www.sidneyrigdon.com/dbroadhu/OH/miscohio.htm#021531>

³ See also:

<http://www.mun.ca/rels/restmov/texts/acampbell/mh1844/MTTBOM.HTM>

⁴ See corrected typescript of Wright's letter prepared by Dale Broadhurst in 2001
at: <http://solomonspalding.com/SRP/saga2/Ashtab3.htm#1833text>

⁵ The eight affidavits collected by Hurlbut and published by Howe (1834) are
available on-line at: <http://www.mormonstudies.com/witness.htm>. A Mormon
response to the Spalding-Rigdon Theory is available at:

<http://farms.byu.edu/display.php?table=review&id=584>

⁶ Solomon Spalding's wife and daughter were both named Matilda. For a
statement from Matilda Spalding Davidson, Solomon's widow, see her letter
to the editor of the Boston Recorder April 19, 1839, at

<http://www.solomonspalding.com/docs1/1897spld.htm>. For statements from

Matilda Spalding McKinstry, Solomon's daughter, see her interview with
Jesse Haven, 1839, and a letter from John Haven, which appeared in the

Quincy Whig, published by Benjamin Winchester in *The Origin of the
Spaulding Story, Concerning the Manuscript Found*, 1840. Copies of both

are available at: <http://www.mormonstudies.com/matilda2.htm>. For a

statement from Robert Patterson, owner of the Pittsburgh publishing shop

where Spalding allegedly took his manuscript, see the statement dated April 2, 1842, to Rev. Samuel Williams in: *Mormonism Exposed*. 1842. See also Samuel Williams's self-published pamphlet (reprinted in the *Baptist Home Mission Monthly*, May 1883) available at

<http://www.solomonspalding.com/docs/1842Wilm.htm#pg16b>

For statements from others who knew Solomon Spalding, see Abner Jackson's statement in Canton, OH, December 20, 1880, published in *The Daily Evening Reporter* (Washington, PA) Vol. 4, Jan 7, 1881. Jackson's father had business dealings with Spalding. His statement is available on-line at

<http://www.sidneyrigdon.com/dbroadhu/PA/penn1860.htm#010781>. Joseph

Miller knew Spalding in the last years of his life, while he lived in Amity, PA.

Miller made five statements over more than three decades (1869, 1879, 1882, 1885, 1890) available at:

<http://www.sidneyrigdon.com/dbroadhu/PA/penn1860.htm#040869>;

<http://www.sidneyrigdon.com/dbroadhu/PA/penn1860.htm#020579>;

<http://www.sidneyrigdon.com/dbroadhu/IA/sain1882.htm#011582>;

<http://www.solomonspalding.com/docs2/1885DicE.htm#pg240b>;

<http://www.solomonspalding.com/docs1/1890GrgD.htm#pg441>

⁷ In mid-December of 1833, Philastus Hurlbut allegedly displayed a copy of Spalding's "Manuscript Found" in or near Kirtland, OH. He was then arrested and incarcerated for threatening the life of Joseph Smith. After his release, Hurlbut never again displayed a copy of "Manuscript Found" or claimed to

possess it. The four witnesses who reported seeing a copy of "Manuscript Found" were C. Downen (Justice of the Peace), James A. Briggs (attorney for Hurlbut), Charles Grover, and Jacob Sherman. Their statements are available at:

<http://solomonspalding.com/SRP/SRP13p2.htm#Refs2>

<http://www.sidneyrigdon.com/dbroadhu/CA/natr1988.htm#120088-1c2>

<http://www.solomonspalding.com/docs/deming.txt>

<http://www.solomonspalding.com/docs/deming.txt>

⁸ A synthesis of historical facts supporting the Spalding-Rigdon theory can be found in Cowdrey, W. L., Davis, H. A. & Vanick, A. (2005) *Who Really Wrote the Book of Mormon?* St. Louis, MO, Concordia Publishing House. See also <http://sidneyrigdon.com/dbroadhu/CA/natruths.htm> and <http://www.sidneyrigdon.com/wht/1891WhtB.htm>.

⁹ The document was found in 1884 by James H. Fairchild and is now stored at the Mudd Library of Oberlin College.

¹⁰ The original basis for this argument is a statement made by James H. Fairchild in the *New York Observer* on February 5, 1885, immediately after his discovery of the Spalding Manuscript in Honolulu, Hawaii: "The theory of the origin of the *Book of Mormon* in the traditional manuscript of Solomon Spaulding will probably have to be relinquished . . . Mr. Rice, myself, and others compared it with the *Book of Mormon*, and could detect no resemblance between the two in general or detail. Some other explanation of the origin of the *Book of Mormon* must be

found, if an explanation is required." Also See:

<http://www.solomonspalding.com/docs2/1886Fair.htm#fairchild1>.

What is often not mentioned is Fairchild's later retraction published by Schroeder (1901): "With regard to the manuscript of Mr. Spaulding now in the Library of Oberlin College, I have never stated, and know of no one who can state, that it is the only manuscript which Spaulding wrote, or that it is certainly the one which has been supposed to be the original of the *Book of Mormon*. The discovery of this Ms. does not prove that there may not have been another, which became the basis of the *Book of Mormon*. The use which has been made of statements emanating from me as implying the contrary of the above is entirely unwarranted. JAMES H. FAIRCHILD." Also, relevant is the final statement of L.L. Rice (1886):

"The Spaulding Manuscript recently discovered in my possession, and published by the Mormons, in no wise determines the question as to the authorship of the Book of Mormon, or of Spaulding's connection with the latter. It shows conclusively that this writing of Spaulding was not the original of the *Book of Mormon* -- nothing more in that regard. It gives the Mormons the advantage of calling upon their opponents to produce or prove that any other Spaulding Manuscript ever existed -- and that is the gist of the whole matter. Until lately I have been of the opinion that there was no tangible evidence that any other production of Solomon Spaulding, bearing upon the question, could be shown as having ever existed. But correspondence and discussions growing out of the publication of this document, have shaken my faith in that belief, and indeed produced quite a change of opinion on that subject."

Rice's statement was published in "The Daily Bulletin", Honolulu, March 11, 1886, Vol IX, No. 1273. It is available on-line at:

<http://www.sidneyrigdon.com/dbroadhu/HI/mischawi.htm#031186>

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

¹¹ Less influential was E. H. Taves's *Trouble enough: Joseph Smith and the Book of Mormon*, Buffalo, N.Y., Prometheus Books, 1984 which relied heavily on the disputed methodologies of A. Q. Morton (see Morton, A. Q. *Literary Detection: How to Prove Authorship and Fraud in Literature and Documents*, New York: Scribner's, 1978).

¹² A further problem with the Larsen study is the observation made by Burrows (1987) in his study of Jane Austen's fiction that authors can create distinct voices within their own prose. Having said that, differences between author "voices" are generally less than that between different authors.

¹³ "The data cited here was extracted from the Early American Fiction (1789-1875) collection published by Chadwyck-Healey (2000).

¹⁴ See also, Hilton, J. L. (1988) "On Verifying Wordprint Studies: *Book of Mormon Authorship*" published by the Foundation for Ancient Research and Mormon Studies, Provo, Utah, 1988.

¹⁵ There are at least three different versions of Hilton's "On Verifying Wordprint Studies: *Book of Mormon Authorship*" available. One version can be found as an undated PDF file in the online *BYU Studies* archive, see <http://byustudies.byu.edu/>. Another copy can be acquired through the Harold B. Lee Library of Brigham Young University under the citation Hilton, J. L. "*Book of Mormon* 'Word Print' measurements using 'wrap-around' block counting," FARMS. FARMS, the Foundation for Ancient Research and Mormon Studies, was recently renamed "The Maxwell Institute for Religious Scholarship" and is located at Brigham Young University. A third source is a book recently

published by FARMS entitled *Book of Mormon Authorship Revisited: The Evidence for Ancient Origins*. See Chapter Nine: "On Verifying Wordprint Studies: *Book of Mormon* Authorship," 225-253.

¹⁶ One could argue that Hilton's decision to remove these biblical sections was reasonable given that the authors he was testing for could not have written these sections. In our analysis we leave these sections in as an additional test of our methodology. A full explanation is provided in the methodology section of this paper.

¹⁷ Joanna Southcott was a late eighteenth century, self-proclaimed, prophet with no connection to the *Book of Mormon*. She was selected by Holmes as a figure similar to Smith who could be used for comparative purposes in his analysis.

¹⁸ See http://www.lightplanet.com/mormons/people/joseph_smith/writings.html

¹⁹ We are aware of objections made to the entire concept of "authorship," objections put forth by theorists such as Stanley Fish and Roland Barthes who have proclaimed that the idea of "authorship" is flawed or "dead." It is not our intention to enter these highly theorized debates but rather to proceed from the assumption that writers do actually write documents and that the documents they write can, with appropriate evidence, be reasonably attributed to their authors. In the case of Joseph Smith, we do not believe that even the small number of letters written in his own hand can be reasonably attributed to him. Moreover, were we to concede the reliability of these few letters, we would still not have enough text to constitute an ample sample of known authorship. This

is a regrettable circumstance, and we do hope that reliable material will be made available in the future for additional testing.

²⁰ Sections 35, 37, 40, 44, 71, 73, 76 and 100 to Smith and Rigdon and sections 6, 7, 13, 18, 24, 26, and 110 to Smith and Cowdery

²¹ In another paper, also published in 1991, Holmes includes the *Book of Abraham* along with the *Doctrine and Covenants* in order to develop Smith's "prophetic" voice. The problems we identify with the *Doctrine and Covenants* are similar in the *Book of Abraham*.

²² Choosing appropriate controls to use in conjunction with the *Book of Mormon* was not a trivial matter. The rationale for our choice is delineated in the section of this paper titled "Source Selection."

²³ More specifically, NSC works by computing a vector of "average word frequencies" for each author and shrinking this vector towards the overall average word frequency vector across all authors in order to reduce the variance and avoid over-fitting the data. A test set sample is then classified by computing its distance to the word frequency vector for each known author, while incorporating knowledge about the variance for each author.

²⁴ See also: <http://www-stat.stanford.edu/~tibs/PAM/index.html>

²⁵ In early July of 1828, Smith lost the first 116 pages of his alleged translation of the inscribed gold plates. Prior to this loss, Smith claimed to have translated the plates by means of the "urim and thummim," a Hebrew instrument of divination that, according to Smith, consisted of a pair of stones fastened to a breastplate

and joined in a form similar to that of a large pair of spectacles. When Smith resumed the alleged translation after loss of the 116 pages, however, he reportedly dictated all of the 582 pages of the 1830 Book of Mormon (including large sections quoted verbatim from the Book of Isaiah) by gazing into a hat in which he held a "seer stone" that allowed him to "see" the words in English and to thus read from the gold plates that were often purportedly hidden to avoid theft . This seer stone was a stone previously used by Smith to look for gold treasures allegedly buried on the land of local farmers, a practice for which Smith was prosecuted (successfully) in a court of law. For a fee, Smith and his associates would dig for treasures at locations identified with the aid of the seer stone. Inevitably, the treasures were cursed and "slippery," preventing their recovery.

²⁶ See for example: <http://scriptures.lds.org/bm/contents>

²⁷ It warrants note that in our cross-validation tests, we did not observe a correlation between whether an author was correctly assigned and the length of a text sample. After NSC and Delta author assignments were made, we further tested the results for any possible correlation between the size of a *Book of Mormon* chapter and the author assigned to that chapter. Again no correlation was observed.

Sample sizes varied as follows:

- Average *Book of Mormon* sample size 1117 (range 91-3752)
- Average Rigdon sample size 4561 (range 226-17797)

- Average Spaulding sample size 2373 (range 777-8515)
- Average Cowdery sample size 1600 (range 200-10712)
- Average Pratt sample size 3024 (range 114-16468)
- Average Longfellow sample size 1354 (range 668-2188)
- Average Barlow sample size 5460 (range 2843-6984)
- Average Isaiah/Malachi sample size 554 (range 134-1131)

²⁸ For example, the Longfellow and Barlow texts frequently open verses with the word "and" as well as frequently stringing together multiple short phrases. Also evident in the three texts is the use of epic or "biblical" language with a propensity for repetition.

²⁹ Fifty novels written by fifteen different authors (twelve male, three female) were selected from the Chadwyck Healey Nineteenth Century American Literature collection. Selection was based solely on publication date (e.g. chronological proximity to the 1830 publication of the *Book of Mormon*). From these texts we extracted word frequency data and employed hierarchical clustering ("hclust" function with complete linkage) to group the texts based on their similarity. The cluster dendrograms produced by R can be found in the online supporting materials at <http://purl.stanford.edu/ir:rs276tc2764>

³⁰ It has been suggested to the authors that the passages from Isaiah and Malachi that appear verbatim in the *Book of Mormon* might be removed from our candidate author samples and perhaps even replaced by similar books from the Bible. Our purpose in keeping these passages was intentional and

meant to serve as an added test of our methodology. Had NSC failed to assign those chapters of the *Book of Mormon* that contain significant borrowings from Isaiah and/or Malachi to our Isaiah/Malachi sample, then we would have had grounds to question the effectiveness of our classifications overall. As it turned out, NSC effectively assigned to Isaiah/Malachi even those chapters where the direct borrowings were subtle.

³¹ All data is available online at <http://purl.stanford.edu/ir:rs276tc2764>

³² In order to avoid the problems that John Burrows (2005) identifies in his study of *Shamela*, our methodology selects for words common to all authors and then, to control for infrequent words that might be common to only one author and the target text, we further winnow the selection to contain only words that appear at a rate of 0.1%. On this point, Hoover (2002) is also instructive. In his analysis of frequent word sequences, Hoover culls from analysis words that are unique to a single text and words with "obviously peculiar distributions" such as those that we found for the words: "god", "ye", "thy", and "behold".

³³ <http://cran.r-project.org/>

³⁴ The twenty-one chapters from the *Book of Mormon* that use the same words in the same sequence as Isaiah or use slightly modified wording are: 1 Nephi 20 (84% identical to Isaiah 48), 1 Nephi 21 (87% identical to Isaiah 49), 2 Nephi 7 (79% identical to Isaiah 50), 2 Nephi 8 (93% identical to Isaiah 51), 2 Nephi 12 (86% identical to Isaiah 2), 2 Nephi 13 (94% identical to Isaiah 3), 2 Nephi 14 (95% identical to Isaiah 4), 2 Nephi 15 (96% identical to Isaiah 5), 2ne16 (96%

identical to Isaiah 6), 2 Nephi 17 (98% identity to Isaiah 7), 2 Nephi 18 (97% identity to Isaiah 8), 2 Nephi 19 (96% identity to Isaiah 9), 2 Nephi 20 (97% identical to Isaiah 10), 2 Nephi 21 (99% identical to Isaiah 11), 2 Nephi 22 (97% identical to Isaiah 12), 2 Nephi 23 (94% identity to Isaiah 13), 2 Nephi 24 (93% identical to Isaiah 14), Mosiah 14 (identical to Isaiah 53), 3 Nephi 22 (identical to Isaiah 54), 3 Nephi 24 (includes all of Malachi 3), and 3 Nephi 25 (identical to Malachi 4).

³⁵ *Ibid.*

³⁶ The thirty-six chapters not conventionally understood as being derived from Isaiah/Malachi that were assigned with a probability above 50% to Isaiah/Malachi are: 2 Nephi 28, 3 Nephi 9, 2 Nephi 26, 2 Nephi 29, 2 Nephi 3, Ether 4, 3 Nephi 21, 2 Nephi 7, 2 Nephi 27, Helaman 13, 2 Nephi 30, 2 Nephi 4, 3 Nephi 30, 3 Nephi 16, 3 Nephi 20, Mosiah 10, Mosiah 17, Mosiah 20, Jacob 5, Mormon 8, Ether 2, 3 Nephi 27, 2 Nephi 10, 3 Nephi 14, 1 Nephi 12, Mosiah 12, 2 Nephi 9, Jacob 3, Ether 3, 3 Nephi 11, 1 Nephi 2, 1 Nephi 7, Mosiah 24, Helaman 14, Mosiah 3, 1 Nephi 13.

³⁷ 2 Nephi 9, 2 Nephi 27, 2 Nephi 28, and 2 Nephi 30 make heavy use of Isaiah's phraseology, such as "Holy One of Israel" "Holy One of Israel" appears twenty-seven times in Isaiah, but just five times in the rest of the Bible. 2 Nephi 10, 2 Nephi 26, 2 Nephi 27, 2 Nephi 28, 2 Nephi 29, 2 Nephi 30, Jacob 5, 3 Nephi 20, 3 Nephi 21 are prophetic chapters dealing with the theme of restoration and expounding on Isaiah. Excerpts from Isaiah are found in 1 Nephi 13, 2 Nephi

26, 2 Nephi 27, 2 Nephi 30, Mosiah 12, 3 Nephi 16, 3 Nephi 20, 3 Nephi 21, and Mormon 8. Verse 20 in 3 Nephi 20 and verse 23 in Mormon 8 both include admonitions to remember or search the words of Isaiah. 3 Nephi 14 is a chapter borrowed from the New Testament (Matthew 7).

³⁸ Ethan Smith was pastor of the Congregationalist Church attended by the family of Oliver Cowdery.

³⁹ According to the official Mormon Church account, Smith received the gold plates upon which the *Book of Mormon* was inscribed from an angel on September 22, 1827. He is said to have begun his translation sometime between December 1827 and February 1828. In mid-June of 1828, the first 116 pages of the document were lost. Shortly after completing the first 116 pages of the document, Smith's scribe Martin Harris took the document home to show his wife. Stories differ as to whether the pages were lost, stolen, or destroyed by Harris's wife.

⁴⁰ The bar charts provided in panel B in Figures 5-10 were inspired by Broadhurst and can be compared to his charts available online. The attributions made by Dale Broadhurst and a detailed analysis are available at <http://solomonspalding.com/SRP/MEDIA/SRPpap16.htm#Alma> and <http://solomonspalding.com/SRP/SRPpap10.htm>

⁴¹ See: <http://www.sidneyrigdon.com/vern/vernP0.htm#pg03>

⁴² 1 Nephi 6, 2 Nephi 32, Mosiah 2, Alma 5, Alma 7, Alma 9, Alma 26, Alma 29, Alma 32, Alma 33, Alma 36, Alma 38, Alma 39, Alma 40, Alma 54, Alma 60, Alma 61, 3 Nephi 12, 3 Nephi 13, and Moroni 1.

⁴³ Cowdery was Smith's primary scribe from April 7 to June 2, 1829

⁴⁴ NSC assigned Alma chapter thirty-six to Oliver Cowdery. This chapter is a chiasm: an inverted parallel literary form. According to B. F. Edwards (2004) the pattern of the chiasm found in Alma thirty-six "establishes with 99.98 percent certainty that this chiasm occurred in this book by design and rules out the hypothesis that it occurred by chance." According to J. W. Welch (2003), publications in the New England area describing use of chiasmus as a Biblical literary form were "available for purchase in bookshops or from traveling salesmen" in 1825. During that time period, Oliver Cowdery worked as a traveling salesman, selling books and pamphlets (see Cowdery et al., 2005).

⁴⁵ According to *History of the Church* and section eight of the *Doctrine and Covenants* (another book of Mormon scripture), Cowdery attempted to translate a part of the *Book of Mormon*, but met with limited success. Both *History of the Church of Jesus Christ of Latter-day Saints* and the *Doctrine and Covenants* are credited to Smith, but the extent of Smith's actual contribution is unknown. Many historical accounts and revelations attributed to Smith were changed after-the-fact by others and/or co-produced with others, including Rigdon and Cowdery.

⁴⁶ 1 Nephi 4, 1 Nephi 5, 1 Nephi 11, 1 Nephi 16, 1 Nephi 18, Mosiah 9, Mormon 2, Moroni 2, and Moroni 3.

⁴⁷ See: <http://www.solomonspalding.com/docs/1901schr.htm> and <http://solomonspalding.com/docs2/1914Shk1.htm#pgvii>

⁴⁸ In Bainbridge Rigdon reportedly became involved in what appears to be “automatic writing”: using a séance-like process to create the *Book of Mormon*. A description of that process is given in a letter to the editor titled “The Mormon Bible” which appeared in the *New Northwest* on September 9, 1880. The letter reads: “We are in receipt of a letter from Mr. O. P. Henry, an Astoria subscriber, who says, in reference to an article in the *Oregonian* of recent date concerning the origin of the Mormon Bible, that his mother, who is yet alive, lived in the family of Sidney Rigdon for several years prior to her marriage in 1827; that there was in the family what is now called a “writing medium,” also several others in adjacent places, and the Mormon Bible was written by two or three different persons by an automatic power which they believed was inspiration direct from God, the same as produced the original Jewish Bible and Christian New Testament. Mr. H. believes that Sidney Rigdon furnished Joseph Smith with these manuscripts, and that the story of the “hieroglyphics” was a fabrication to make the credulous take hold of the mystery; that Rigdon, having learned, beyond a doubt, that the so-called dead could communicate to the living, considered himself duly authorized by Jehovah to found a new church, under a divine guidance similar to that of Confucius, Moses, Jesus,

Mohammed, Swedenborg, Calvin, Luther or Wesley, all of whom believed in and taught the ministration of spirits. The *New Northwest* gives place to Mr. Henry's idea as a matter of general interest. The public will, of course, make its own comments and draw its own conclusions." See:

<http://www.sidneyrigdon.com/dbroadhu/NW/miscnw04.htm#081680>. Dale

Broadhurst has confirmed several aspects of the above account, and compiled additional historical evidence pointing to Bainbridge as the likely location for production of the 1827 version of the *Book of Mormon*. See:

<http://sidneyrigdon.com/books/Hnry1942.htm> and

<http://www.sidneyrigdon.com/books/Brew1945.htm>

⁴⁹ See <http://solomonspalding.com/docs/1882PatA.htm>

⁵⁰ Several thematic similarities to the *Book of Mormon* have been suggested by Holley (1989), Broadhurst (<http://www.sidneyrigdon.com/>) and Chandler (<http://mormonstudies.com/>). Tom Donofrio (see <http://www.mormonstudies.com/early1.htm>) has shown that Spalding's Oberlin Manuscript and *The Book of Mormon* both contain phrases borrowed from David Ramsay (1749-1815), a friend and biographer of George Washington and author of *History of the American Revolution*. They also contain phrases from Mercy Otis Warren (1728-1814), author of *History of the Rise, Progress and Termination of the American Revolution* (1805). These borrowed phrases are concentrated in sections on war within the Book of Alma, where the

Spalding signal that NSC detected is most pronounced (Figure 9A and 9B, chapters 138-143 in particular).

See: <http://www.mormonstudies.com/early1.htm>

http://www.postmormon.org/exp_e/index.php/magazine/pmm_article_full_text/2

11

⁵¹ For a description of the Joseph Smith papers project, see:

<http://www.josephsmithpapers.net>

⁵² Since Van Wagoner's 1994 biography of Sidney Rigdon, Mormon history researchers have become increasingly aware of the pivotal role Rigdon played in the emergence of Mormonism. Recently, for example, Reynolds (2005) concluded that Rigdon was the likely author of "The Lectures on Faith," a series of seven lectures previously attributed to Joseph Smith. These lectures played a key role in the development of early Mormon theology.

References

"Mormonism--or Grand Pugilistic Debate" (1831). *Cleveland Advertiser*, Volume 1, Number 5, Tuesday, February 15.

Barlow, J. (1825). *The Columbiad*, Washington [D.C.], J. Milligan.

Brodie, F. (1971). *No Man Knows My History; the Life of Joseph Smith, the Mormon Prophet*, New York, Knopf.

Brodie, F. M. (1945). *No Man Knows My History; the Life of Joseph Smith, the Mormon Prophet*, New York, A. A. Knopf.

Burrows, J. F. (1987). *Computation into Criticism: a Study of Jane Austen's Novels and an Experiment in Method*. Oxford, Clarendon Press.

Burrows, J. F. (2002). 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing*, 17, 267-87.

Burrows, J. F. (2003). Questions of Authorship and Beyond. *Computers and the Humanities*, 37, 5-32.

Burrows, J. F. (2005). Who Wrote *Shamela*? Verifying the Authorship of a Parodic Text. *Literary and Linguistic Computing*, 20, 437-450.

Bushman, R. L. (2005). *Joseph Smith: Rough Stone Rolling*, New York, Alfred A. Knopf.

Campbell, A. (1832). *Delusions: An Analysis of the Book of Mormon, With an Examination of its Internal and External Evidences, and a Refutation of its Pretences to Divine Authority...*, Boston, MA, Benjamin Greene.

Campbell, A. (1844). Mistakes Touching the Book of Mormon. *The Millennial Harbinger*, 1, 38-42.

- Chadwyck-Healey (2000). *Early Early American Fiction 1789–1875*. ProQuest Information and Learning and the University of Virginia Library.
- Cowdrey, W. L., Davis, H. A. & Vanick, A. (2005). *Who Really Wrote the Book of Mormon?*, St. Louis, MO, Concordia Pub. House.
- Croft, J. D. (1981). Book of Mormon 'Wordprints' Reexamined. *Sunstone* 6, 15-21.
- Deming, A B., ed. (1888). *Naked Truths About Mormonism*. Oakland, California: Deming and Company.
- Eaton, H (1882). Origin of Mormonism. In: *Handbook of Mormonism*, Salt Lake City, UT, Handbook Pub.
- Edwards, B. F. Edwards W. F. (2004). Does chiasmus appear in the Book of Mormon by chance? *BYU Studies* 43, 103-130.
- Farmer, S. (2006). Neurobiology, Stratified Texts, and the Evolution of Thought: From Myths to Religions and Philosophies. Presented at the Harvard and Peking University International Conference on Comparative Mythology, Beijing, China, 11-13 May. See <http://www.safarmer.com/Farmer.Beijingpaper2.pdf>
- Hilton, J. L. (1988). On Verifying Wordprint Studies: Book of Mormon Authorship. Foundation for Ancient Research and Mormon Studies, Provo, Utah 1988.
- Holley, V. (1989). *Book of Mormon Authorship: A Closer Look*, UT, Self-published.
- Holmes, D. I. (1985). The Analysis of Literary Style—A Review. *Journal of the Royal Statistical Society, Series A (General)*, 148, 328-341.
- DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

Holmes, D. I. (1991a). A Multivariate Technique for Authorship Attribution and Its Application to the Analysis of Mormon Scripture and Related Texts.

History and Computing, 3, 12-22.

Holmes, D. I. (1991b). Vocabulary Richness and the Prophetic Voice. *Literary and Linguistic Computing*, 6, 259-268.

Holmes, D. I. (1992). A Stylometric Analysis of Mormon Scripture and Related Texts. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 155, 91-120.

Hoover, D. L. (2002). Frequent Word Sequences and Statistical Stylistics.

Literary and Linguistic Computing, 17, 157-180.

Hoover, D. L. (2004a). Delta Prime? *Literary and Linguistic Computing*, 19, 477-493.

Hoover, D. L. (2004b). Testing Burrow's Delta. *Literary and Linguistic Computing*, 19, 453-475.

Howe, E. D. (1834, 1977). *Mormonism Unveiled: or, A faithful account of that singular imposition and delusion, from its rise to the present time: with sketches of the characters of its propagators, and a full detail of the manner in which the famous Golden Bible was brought before the world*, New York, AMS Press.

Larsen, W. A., Rencher, A. C., Layton, T (1980). Who Wrote the Book of Mormon? An Analysis of Word-prints. *BYU Studies*, 20, 225-51.

Longfellow, H. W. (1855). *The Song of Hiawatha*, Boston, Ticknor and Fields.

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491

- Madsen, B. D. (1985). *Studies of the Book of Mormon*, Urbana, University of Illinois Press.
- Marquardt, H. M. (2000). *Literary Dependence in the Book of Mormon: Two Studies*. Institute for Religious Research. <http://www.irr.org/mit/literary-dependence-bom-1a.html>
- Morton, A. Q. (1978). *Literary Detection: How to Prove Authorship and Fraud in Literature and Documents*, New York, Scribner's.
- Palmer, G. H. (2002). *An Insider's View of Mormon Origins*, Salt Lake City, Signature Books.
- Persuitte, D. (2000). *Joseph Smith and the Origins of the Book of Mormon*, Jefferson, North Carolina, McFarland & Co. Publishers.
- Priest, J. (1825). *The Wonders of Nature and Providence, Displayed. Compiled from authentic sources, both ancient and modern, giving an account of various and strange phenomena existing in nature, of travels, adventures, singular providences, &c*, Albany, J. Priest.
- Reynolds, N. B. (2005). The Case for Sidney Rigdon as Author of the Lectures on Faith. *Mormon History*, 32, 1-41.
- Rice, L. "Mormonism." *The Daily Bulletin*, Honolulu, HI: Volume 9, Number 1273, March 11, 1886.
- Riley, I. W. (1902). *The Founder of Mormonism: A Psychological Study of Joseph Smith, Jr.*, New York, Dodd, Mead, and Co.

- Schroeder, T. A. (1901). *The Origin of the Book of Mormon, Re-Examined in Its Relation to Spaulding's "Manuscript Found"*. Salt Lake City, UT: Salt Lake Ministerial Association.
- Shook, C. A. (1914). *True Origin of The Book of Mormon*, Cincinnati, OH, Standard Publishing Co.
- Smith, E. (1825). *View of the Hebrews*, Poultney, VT, Smith & Shute.
- Smith, J. & Jessee, D. C. (2002). *The Personal Writings of Joseph Smith*, Salt Lake City, Brigham Young University Press, Deseret Books.
- Spaulding, S. (1810). *The Solomon Spaulding Manuscript*. Oberlin College Library.
- Tanner, J. A. S. T. (1998). *Joseph Smith's Plagiarism of the Bible*, Salt Lake City, Utah Lighthouse Ministry.
- Taves, E. H. (1984). *Trouble enough: Joseph Smith and the Book of Mormon*, Buffalo, N.Y., Prometheus Books.
- Tibshirani, R., Trevor Hastie, Balasubramanian Narasimhan and Gilbert Chu (2003). Class Prediction by Nearest Shrunken Centroids, with Applications to DNA Microarrays. *Statistical Science*, 18, 104-117.
- Tibshirani, R., Trevor Hastie, Balasubramanian Narasimhan, and Gilbert Chu (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences*, 99, 6567-6572.
- Vogel, D. (2004). *Joseph Smith: The Making of a Prophet*, Salt Lake City, Signature Books.

- Walters, W. P. (1990). *The Use of the Old Testament in the Book of Mormon*.
Utah Lighthouse Ministry, Salt Lake City, Covenant Theological Seminary.
- Welch, J. W. (2003). How Much Was Known About Chiasmus in 1829 When the
Book of Mormon Was Translated? *FARMS Review*, 15, 47-80.
- Whitsitt, W. H. (1885, 1988). *Sidney Rigdon, The Real Founder of Mormonism*.
Salt Lake City, UT: Metamorphosis Pub.
- Williams, S. (1842). *Mormonism Exposed*, Pittsburgh, PA, Self-published.
- Wright, A. (1833). Draft Letter to Josiah Jones, Grandison Newell or Warren
Corning (uncertain). Solomon Spalding Miscellaneous Papers. New York,
New York Public Library.

Appendix A: Source Materials

The *Book of Mormon*: For analysis, we encoded into XML an electronic copy of the 1830 text version of the *Book of Mormon* obtained from Mr. Ronald Dawbarn and modified to include modern LDS chapter divisions. For our version we excluded Book prefaces. A copy of our text can be found online at <http://www.stanford.edu/~mjockers/bomData/>

Old Testament: Because sections of the *Book of Mormon* are known to be borrowed or derived specifically from the Books of Isaiah and Malachi, we used the entire Book of Isaiah and the first four chapters from Malachi in the analysis.

Solomon Spalding: The "Oberlin Manuscript," written sometime between 1810 and 1812 and published posthumously in 1910, is the only Spalding text available today. The handwritten original is stored in the Mudd Library of Oberlin College, OH. The manuscript has been referred to by many names including the "Oberlin Manuscript," "Manuscript Story," "Manuscript Story–Conneaut Creek," the "Roman Story," "Fabius Story," "Conneaut Story," and the "Honolulu" manuscript. Our electronic version of the document was derived from the document originally published at the Millennial Star office, Liverpool, England. We include the letter appended to the Oberlin Manuscript, which is written in the same hand as the Oberlin manuscript. Our XML source file was derived from the HTML version available at <http://www.mormonstudies.com/spaldg1.htm>

Sidney Rigdon: An electronic copy of the *Book of the Revelations of Jesus*

Christ to the Children of Zion (written 1864) was provided by Mr. David Marshall who transcribed the hand-written document, which is available in *Copying Books A & B* of the Stephen Post Collection at the University of Utah. We removed from this document specific citations of Old and New Testament scripture and the revelations to Phebe Rigdon (Sidney Rigdon's wife). We also utilized *Millenium* (published December 1833 to May 1835), a collection of 14 articles published in the newspapers of the Latter Day Saints. This collection is available on-line at: <http://www.sidneyrigdon.com/RigWrit/RigWrit3.htm/>. Passages from the Old and New Testament were removed.

Oliver Cowdery: A series of writings from the Latter Day Saints' *Messenger and Advocate*, were obtained from: <http://www.boap.org/LDS/Early-Saints/Cowdery-hist.html>. Passages from the Old and New Testament were removed, as were other passages not attributable to Cowdery.

Parley P. Pratt: Four texts from Pratt were used in our analysis. Passages from the Old and New Testament were removed from all, as were other passages not attributable to Pratt. The texts are as follows:

----. *The Autobiography of Parley P. Pratt*. (1938, 1985) Deseret Book Company, Salt Lake City, Utah. Edited by Parley P. Pratt, Jr.
<http://jared.pratt-family.org/old/histories/ppp-autobiography/>

----. *Mormonism Unveiled . . . Zion's Watchman Unmasked, And Its Editor, Mr. L. R. Sunderland: Exposed: Truth Vindicated: The*

Devil Mad, and Priestcraft in Danger! (Second Edition: Nyc, 1838).

<http://solomonspalding.com/docs/prt1838b.htm#pagetop>.

----. *The Angel of the Prairies: A Dream of the Future*. Abinadi Pratt,

Publisher. Salt Lake City, January, 1880.

[http://mlmb.byu.edu/Anthology/Pratt-](http://mlmb.byu.edu/Anthology/Pratt-Angel%20of%20the%20Prairies.htm)

[Angel%20of%20the%20Prairies.htm](http://mlmb.byu.edu/Anthology/Pratt-Angel%20of%20the%20Prairies.htm)

----. *Plain Facts Showing The Falsehood And Folly Of The Rev.*

C. S. Bush, (A Church Minister Of The Parish Of Peover,) Being A

Reply To His Tract Against The Latter-Day Saints. (Manchester,

UK: 1840) New York, Nov. 27th, 1839.

<http://www.solomonspalding.com/docs/prt1840d.htm>.

Appendix B

110 Words Used in NSC Classification for the seven-author case.

1	a	38	he	75	said
2	after	39	her	76	say
3	again	40	him	77	shall
4	against	41	his	78	should
5	all	42	i	79	so
6	among	43	if	80	son
7	an	44	in	81	that
8	and	45	into	82	the
9	are	46	is	83	their
10	as	47	it	84	them
11	at	48	king	85	then
12	away	49	know	86	there
13	be	50	land	87	therefore
14	because	51	made	88	these
15	been	52	man	89	they
16	before	53	many	90	things
17	but	54	may	91	this
18	by	55	me	92	those
19	came	56	men	93	thus
20	children	57	might	94	time
21	come	58	more	95	to
22	day	59	my	96	up
23	did	60	name	97	upon
24	do	61	no	98	us
25	down	62	not	99	was
26	earth	63	now	100	we
27	even	64	o	101	were
28	every	65	of	102	when
29	father	66	on	103	which
30	for	67	one	104	who
31	forth	68	or	105	will
32	from	69	our	106	with
33	go	70	out	107	words
34	great	71	over	108	would
35	had	72	pass	109	you
36	hand	73	people	110	your
37	have	74	power		

108 Words Used in NSC Classification for the five-author case

1	a	37	her	74	say
2	after	38	him	75	shall
3	again	39	his	76	should
4	against	40	i	77	so
5	all	41	if	78	son
6	among	42	in	79	that
7	an	43	into	80	the
8	and	44	is	81	their
9	are	45	it	82	them
10	as	46	king	83	then
11	at	47	know	84	there
12	away	48	land	85	therefore
13	be	49	made	86	these
14	because	50	man	87	they
15	been	51	many	88	things
16	before	52	may	89	this
17	but	53	me	90	those
18	by	54	men	91	thus
19	came	55	might	92	time
20	children	56	more	93	to
21	come	57	my	94	up
22	day	58	name	95	upon
23	do	59	no	96	us
24	down	60	not	97	was
25	earth	61	now	98	we
26	even	62	o	99	were
27	every	63	of	100	when
28	for.	64	on	101	which
29	forth	65	one	102	who
30	from	66	or	103	will
31	go	67	our	104	with
32	great	68	out	105	words
33	had	69	over	106	would
34	hand	70	pass	107	you
35	have	71	people	108	your
36	he	72	power		
		73	said		

DRAFT PREPRINT OF "Reassessing Authorship of the *Book of Mormon* Using Delta and Nearest Shrunken Centroid Classification." *Literary and Linguistic Computing*, December 2008; 23: 465 - 491