Measuring shared authorship in the Federalist Papers, Shakespeare and the Bible

Matthew D. Schwartz
Department of Physics, Harvard University

October 1, 2026

This work is preliminary. Some content remains AI-generated and is still being validated.

Abstract

When primary historical sources leave a text's authorship unresolved, it may still be decidable from statistics of the text itself. Since Mosteller and Wallace's study of the Federalist, function-word counts have mostly served to assign texts to single authors or locate style changes. Hypotheses of mixed or multiple authorship seldom receive probabilities of their own. Treating a disputed text as a statistical mixture gives each hypothesis its own Bayesian evidence, the probability of the observed counts under that hypothesis. We compute every evidence in closed form, many in exact arithmetic, and measure detection power by inserting a second author's text into solo works. The two-author analyses are repeated under a second model in which authors' rates vary between texts. Across the Federalist, the Shakespeare canon, and the Bible our results mostly confirm the scholarly consensus and make it precise; a few do not. Hamilton and Madison's joint papers are Madison-dominant under every word list and model. Six Shakespeare plays show a second hand, four under both models; in *Henry VIII* it is closer to Middleton's profile than Fletcher's, and four other plays, long enough to test, show none. Isaiah divides after chapter 39 in two manuscripts a millennium apart, and neither supports a third Isaiah at the power the later chapters allow. In the Pentateuch, Priestly material separates even within books, the Deuteronomic separation coincides with the book of Deuteronomy, and sources J and E cannot be told apart. The Bible's books fall into roughly a dozen to thirty distinguishable profiles, depending on prior and vocabulary, several spanning canonical sections.

Keywords: authorship attribution; stylometry; Bayesian model comparison; marginal likelihood; mixture models; changepoint analysis; Federalist Papers; Shakespeare; Hebrew Bible.

1 Introduction

Who wrote a text is ordinarily settled by evidence outside it: a title page, a printer's register, a letter, the word of a contemporary. When such evidence is missing or contradictory, the case rests on the language of the text itself, on whether habits of wording can tell one writer from another. The best-known disputes of this kind concern the *Federalist*, the Shakespeare canon, and the Bible. Hamilton and Madison left rival lists claiming some of the same *Federalist* essays and acknowledged writing three of them together. The share of each man in those three, and the authorship of the twelve essays that both claimed, have been debated since. Several plays printed as Shakespeare's have long been suspected of containing another dramatist's work. For Isaiah and the five books of

Moses, critics have proposed anything from one writer to many. Statistical attribution was built on these cases and is still judged against them.

Attempts to decide authorship by counting features of the text go back more than a century, to Mendenhall (1887) on word lengths and Yule (1939, 1944) on sentence length and vocabulary. Mosteller and Wallace (1964) put the comparison on a probabilistic footing. They modeled the rates of function words (closed-class words such as prepositions and conjunctions, whose use varies little with subject matter) in the undisputed *Federalist* essays and computed posterior odds that assigned all twelve disputed essays to Madison. Most later work frames attribution as classification, the choice of one author per text from a list of candidates (Burrows, 2002; Stamatatos, 2009; Koppel et al., 2009). Word-rate tests on plays and scenes underlie much of the current account of Shakespeare's collaborators (Craig and Kinney, 2009; Taylor and Egan, 2017). For a text with more than one hand, existing methods locate the change of style or divide the text among a given number of authors (Koppel et al., 2011; Plecháč, 2021); biblical studies test whether the parts of a source-critical partition differ (Radday, 1970; Dershowitz et al., 2015; Yoffe et al., 2023; Faigenbaum-Golovin et al., 2024). Puig et al. (2016) and Font et al. (2016) used closed-form probabilities of the word counts, with the unknown rates averaged over, to choose among candidate authors and, at one fixed grouping of acts or chapters, among numbers of style groups. Literary and statistical studies have converged on a broadly shared picture: the joint *Federalist* papers mainly Madison's, a second dramatist in several plays, a second prophet in the later chapters of Isaiah, and a Priestly source distinct from the rest of the Pentateuch. Yet the composite hypothesis itself (that a paper is one-third Hamilton, or that a book had four sources) is seldom given a probability of its own.

In this paper we treat a disputed text as a statistical mixture. Its function-word counts are modeled as independent draws from one author's rates, from a blend of two authors' rates in an unknown proportion, or from an unknown number of sources, each producing a contiguous run of chapters. Every proposed division of a corpus is then a hypothesis of single, mixed, or multiple authorship. We measure its support by its Bayesian evidence (the marginal likelihood), the probability of the observed counts with the unknown rates, proportions, and boundaries integrated out. Every such evidence, including the integral over the mixing proportion and the sum over all contiguous segmentations and their number, has a closed form or is a finite sum of closed forms, derived in a companion paper (Schwartz, 2026) that builds on Lin et al. (2009). We evaluate the evidences directly rather than by the usual posterior samplers or asymptotic approximations (Evans et al., 1989; Hjort, 2007), many in exact rational arithmetic, so that each hypothesis receives a probability of its own, free of simulation error, across whole canons. We measure detection power by inserting text of known authorship into single-author works. We also repeat the two-author analyses under a hierarchical model, evaluated in the same way, in which each text's rates vary about its author's profile as Mosteller and Wallace (1964) found. Content, theology, and meter, on which much of the traditional argument rests, do not enter the counts.

The three corpora are the 85 *Federalist* papers, the 41 plays of the Shakespeare canon, and the 66 books of the Bible (Hebrew, Aramaic, Greek). We find that the joint *Federalist* papers are Madison-dominant under every word list, with or without the hierarchical model, and show no Hamilton share beyond that of papers Madison wrote alone. *Federalist* No. 55, which Hamilton claimed, leans toward Madison under the two word lists most accurate on the undisputed papers. Six plays show a second dramatist's hand, and four show none at lengths where a fifth of the play in a second hand would have changed the classification at least eight times in ten; the rest are unresolved. Under the hierarchical model four of the six detections remain, as does one of the

four "none" results. All three parts of *Henry VI* fit a Shakespeare–Marlowe blend better than either profile alone, as does a play of Marlowe's own, so the preference does not by itself show collaboration. In a medieval codex and the Great Isaiah Scroll alike, Isaiah divides after its thirty-ninth chapter, where a second prophet has been placed since Döderlein (1789). The function words give no support to a third (Duhm, 1892), though in the later chapters alone a further change could be missed. In the Pentateuch, Priestly material separates from the rest on function words even within books and genres, the Deuteronomic separation coincides with the book of Deuteronomy, and J and E cannot be told apart. Grouped by function-word rates across the canon, the books fall into roughly a dozen to thirty distinguishable profiles, depending on the prior and the vocabulary compared, several spanning canonical sections; these are counts of profiles, not of authors or sources.

Few of our results depart from positions scholars already hold. The second hand in *Henry VIII* lies closer to Middleton's function-word profile than to Fletcher's under both models, and in *Titus Andronicus*, under word independence, Peele's profile is decisive in the third act, whereas the scholarly division gives him the first. Ezekiel, often read as one prophet's book, divides more sharply than any other single-language book, between the oracles and the temple vision. Acts separates from Luke, usually ascribed to the same author. These departures may reflect genre or period style as much as a second hand. Section 2 reviews external and internal evidence and stylometry, Section 3 states the models and checks them on four books of known structure, and the three corpora follow in turn.

2 Authorship disputes and word counts

Authorship disputes are argued from evidence outside the text and inside it, and for more than a century also from counts of its words.

2.1 External and internal evidence

External evidence comes from outside the work: title pages and dedications, entries in the Stationers' Register, the office-book of the Master of the Revels, publishers' records and account books, correspondence, and contemporary allusion (Love, 2002; Schoenbaum, 1966). Internal evidence is read out of the work itself: parallel passages, verbal echoes, meter, style. In Schoenbaum's ranking, external evidence "may and often does provide incontestable proof"; internal evidence "can only support hypotheses or corroborate external evidence"; and the scholar who argues from internal evidence alone "assumes the full burden of proof" (Schoenbaum, 1966).

External evidence, however, is scarce and sometimes wrong: title pages credited single authors for plays that account books show were collaborative, and the office-book of the Master of the Revels has perished in the original. Most live disputes are therefore fought on internal evidence, and Schoenbaum (1966) and Love (2002) cataloged how such arguments fail. Plagiarism or coincidence may explain a parallel passage as well as shared authorship does. Reasoning turns circular when a passage assigned on a hunch becomes the standard against which the next assignment is tested, or, as Love (2002) quotes David Lake, "without some unquestioned attributions no other attributions can be questioned."

The standards of proof are borrowed from the courtroom. Love (2002) writes that the criteria for accepting an attribution as proven "have traditionally been based on legal models for the evaluation of evidence." The probabilities at stake are argued in words and seldom computed.

2.2 Stylometry from Mendenhall to Burrows

The idea that counting could settle authorship questions is Victorian. Augustus De Morgan proposed in an 1851 letter that the average number of letters per word would tell whether the Epistle to the Hebrews was Paul's (De Morgan, 1882). The physicist T. C. Mendenhall (1887) then made such counts, plotted a word-length distribution (one "characteristic curve" per author), and compared the curves by eye. By 1901, with a Boston patron paying for a count aimed at the Bacon–Shakespeare controversy, Mendenhall had two hired counters register nearly two million words on a button-operated machine (Mendenhall, 1901). He reported that Marlowe agreed with Shakespeare "about as well as Shakespeare agrees with himself."

The word "stylometry" is Wincenty Lutosławski's, whose book on the chronology of Plato's dialogues (Lutosławski, 1897) compared some five hundred stylistic peculiarities across twenty-two dialogues and predicted "a new science of stylometry subsidiary to historical research, like palaeography." Yule (1939, 1944) brought sentence-length distributions to disputed authorship and vocabulary statistics to the *Imitatio Christi*. Ellegård (1962) applied machine arithmetic to hand counts of word use in the Junius letters. For Mendenhall and Yule the method was a summary statistic compared across texts, with no probability attached to any hypothesis.

The study of the disputed Federalist papers by Mosteller and Wallace (1963, 1964), which Stamatatos (2009) still calls seminal, did attach probabilities to hypotheses. They used function-word rates to discriminate Hamilton from Madison, modeled the overdispersion of the counts with a negative binomial distribution, and computed posterior odds for each disputed paper. Later methods, such as Burrows's Delta (Burrows, 2002), treat attribution as a choice among candidate authors. Stamatatos (2009) describes the problem as "a multi-class single-label text categorization task," and the machine-learning literature frames it the same way (Koppel et al., 2009; Jockers and Witten, 2010). Bayes factors between single candidate authors have also been computed: Hamilton, Madison, or an unlisted third for a disputed Federalist paper (Puig et al., 2016), Molière or Corneille for a play (Bozza et al., 2025).

The cumulative-sum technique served defense teams in British and Irish cases in 1991 and 1992, before published tests found it unreliable (Grant, 2022; Hardcastle, 1997). In 2013, by contrast, word lengths, character strings, word pairs, and the hundred most frequent words pointed to Rowling among four candidates for a pseudonymous novel (Juola, 2015).

2.3 Mixed and multiple authorship

Unsupervised methods decompose a document into authorial components when their number is supplied (Koppel et al., 2011; Akiva and Koppel, 2013). Rolling attribution locates changes of author (Plecháč, 2021); systems in the PAN style-change detection tasks predict how many authors a document has (Zangerle et al., 2019); and Lennon–McCartney songs have been split into single-author fragments and classified (Glickman et al., 2019). Author–topic models assign the words of a multi-author document among its listed authors and their topics (Rosen-Zvi et al., 2004; Seroussi et al., 2014).

The work nearest ours treats a text's word counts as multinomial and authorship as Bayesian model choice. Gill and Swartz (2011), extending Gill et al. (2007), place a Dirichlet-process prior on the texts' rate vectors, so that the number of authorial groups is not fixed. They report the posterior, sampled by Markov chain Monte Carlo, as the probability that two texts share an author. Girón et al. (2005), Puig et al. (2015), and Font et al. (2016) model the chapters of a book, or the

acts of Shakespeare's plays, as a changepoint sequence or a finite mixture of multinomials. They compare one, two, or more style groups by Dirichlet–multinomial marginal likelihoods evaluated at an estimated grouping, and Puig et al. (2016) choose among candidate authors by the same marginal likelihoods in closed form.

In biblical studies Wickmann (1968), Radday (1970), and Radday and Shore (1985) asked whether two given texts share an author. Recent work tests proposed source partitions (the Priestly/non-Priestly division among them) against word statistics (Dershowitz et al., 2015; Yoffe et al., 2023; Faigenbaum-Golovin et al., 2024). The methods reviewed here locate a change of author, assign a text or its parts to authors, test a given partition, or count style groups. None reports a posterior distribution over an author's share of a text, and the group counts rest on one estimated grouping of the units or on a Dirichlet-process prior.

3 Texts as mixtures

We reduce each text to its function-word counts. A proposed division of a corpus, into one author's work, a blend of two, or sources laid end to end, is a hypothesis about those counts, weighed against rival divisions by its evidence. The evidence is the probability of the counts with every unknown rate, mixing proportion, and boundary integrated out. Each evidence below has a closed form or is a finite sum of closed forms (Lin et al., 2009; Schwartz, 2026), evaluated without sampling or asymptotic approximation.

3.1 Word counts and the evidence integral

The data for a text are the counts $U = (U_1, \ldots, U_k)$ of k listed function words (articles, prepositions, conjunctions, pronouns, particles). We call the list a word list or channel set and each word (one multinomial category) a channel; "channel" alone also denotes the whole list. For Hebrew, Aramaic, and Greek the list comprises the closed morphological classes of a tagged text, pronominal suffixes included (Appendix E). Units are papers, plays, acts, scenes, chapters, or verse blocks (runs of consecutive verses of at least 400 running-text tokens). Under a hypothesis H with parameters θ, the probability of the list-word sequence and the evidence of H are

$$P(U \mid \theta) \;=\; \prod_{v=1}^{k} p_v(\theta)^{U_v}\,, \qquad Z(U) \;=\; \int_{\Theta} \prod_{v=1}^{k} p_v(\theta)^{U_v}\, d\theta\,. \tag{3.1}$$

Here $p_v(\theta)$ is the probability that a token is word v, $d\theta$ is the prior measure on the unknowns Θ, and $Z(U)$ is the marginal likelihood. Tokens are treated as independent draws: the word-count likelihood is a declared iid idealization, which we call the word-independence model, and every Bayes factor below is conditional on it or on its hierarchical extension.

For one homogeneous text with a symmetric Dirichlet prior of parameter α on its rates p, (3.1) becomes

$$Z_\alpha(U) \;=\; \int \prod_{v=1}^{k} p_v^{U_v}\, \mathrm{Dir}(p \mid \alpha)\, dp \;=\; \frac{\Gamma(k\alpha)}{\Gamma(k\alpha + N)} \prod_{v=1}^{k} \frac{\Gamma(\alpha + U_v)}{\Gamma(\alpha)}\,, \qquad N = \sum_{v=1}^{k} U_v\,, \tag{3.2}$$

with N the number of list tokens (Lin et al., 2009; Schwartz, 2026). Puig et al. (2016) used this Dirichlet–multinomial form (Mosimann, 1962) as the marginal likelihood for choosing among

candidate authors, and Font et al. (2016), at one fixed grouping of acts, for choosing among numbers of style groups. We add the integral over the mixing proportion (3.3) and the sums over all contiguous segmentations and their number (3.5). We set α to $1/k$, $1/2$ (Jeffreys, primary), or 1 (uniform) in the changepoint analyses and to 1, $1/2$, or $1/6$ for the Pentateuch. For rational α each factor in (3.2) is a ratio of integers, as are the sums and polynomial integrals below, which we evaluate in exact rational arithmetic, a technique from computer-assisted proofs.

3.2 The two-author blend

For a text that two known authors may share, the hypotheses are that A wrote it, that B did, or that it blends them in unknown proportion. The profiles p_A and p_B are the list-word rates of each author's undisputed texts with add-one smoothing ($\alpha_0 = 1$), held fixed. Under the blend each token is drawn from A's profile with probability w and from B's otherwise, with a uniform $\mathrm{Beta}(1,1)$ prior on w:

$$p_v(w) \;=\; w\,p_{A,v} + (1-w)\,p_{B,v}\,, \qquad Z_{\text{blend}}(U) \;=\; \int_0^1 \prod_{v=1}^{k} \big[\, w\,p_{A,v} + (1-w)\,p_{B,v} \big]^{U_v}\, dw\,. \tag{3.3}$$

The pure hypotheses are $w = 1$ and $w = 0$. The integrand, a polynomial of degree N in w, integrates in closed form (Schwartz, 2026) and, normalized, is the posterior density of w, whose mean and quantiles we report. For the Federalist, A is Hamilton, B is Madison, and the posterior mean of w is the Hamilton share, the expected fraction of a paper's list tokens drawn from Hamilton's profile. For the plays A is Shakespeare and B one of eight other dramatists, Marlowe or Kyd for *Henry VI*, and the best blend is set against the best single profile. A blend can only reweight the profiles supplied: hands outside a declared pool are undetectable by construction, and an unmodeled author's text is assigned to the modeled combination it most resembles.

Mosteller and Wallace (1964) adopted the negative binomial because an author's rate for a word varies from paper to paper more than chance sampling allows (see also Airoldi et al., 2006). In the *hierarchical model* each text d by author a has its own rates,

$$p_{a,d} \;\sim\; \mathrm{Dir}(\kappa_a\, p_a)\,, \tag{3.4}$$

scattered about the profile p_a, with concentration κ_a fitted to the author's undisputed texts by leave-one-out empirical Bayes. A text's counts are then Dirichlet–multinomial (Madsen et al., 2005), as in earlier attribution models (Allison and Guthrie, 2008; Ross, 2020; Ishihara, 2023). The blend mixes $p_{A,d}$ and $p_{B,d}$ token by token, every evidence is again a finite sum in closed form, and $\kappa_a \to \infty$ recovers the word-independence model.

3.3 Sources in sequence

For a book that may combine sources laid end to end, the units are its n chapters or verse blocks in order and a hypothesis divides them into g contiguous segments. We call each segment's source, with its own function-word rates, a contiguous generating process. Thus the changepoint instrument counts processes, never authors, and a seam can mark genre or language rather than hands. Every segment's rates are integrated out by (3.2) and the prior over the $\binom{n-1}{g-1}$ divisions into g segments is uniform. The evidence of g processes, the posterior on g, and the posterior probability of a

boundary after unit t are then

$$Z_g \;=\; \binom{n-1}{g-1}^{-1} \sum_{0=t_0<t_1<\cdots<t_g=n} \prod_{s=1}^{g} Z_\alpha\big(U^{(t_{s-1},\,t_s]}\big)\,, \qquad P(g \mid U) \;=\; \frac{\pi(g)\,Z_g}{\sum_{g'=1}^{6}\pi(g')\,Z_{g'}}\,,$$
$$P(t \mid g, U) \;=\; \frac{1}{\binom{n-1}{g-1}\,Z_g} \sum_{\{t_s\}\ni t} \prod_{s=1}^{g} Z_\alpha\big(U^{(t_{s-1},\,t_s]}\big)\,. \tag{3.5}$$

Here $U^{(a,b]}$ pools units $a+1$ through b, the last sum runs over the divisions with a boundary after unit t, and the prior $\pi(g)$ is uniform on $g \le 6$ (alternatively geometric). The sums are evaluated by the forward-backward recursions of product-partition changepoint models (Barry and Hartigan, 1992, 1993; Liu and Lawrence, 1999; Fearnhead, 2006; Schwartz, 2026). The same construction with whole books as units, grouped without the contiguity constraint, gives the process counts for each canonical section (Schwartz, 2026). The documentary hypothesis posits interleaved sources, which a contiguous segmentation cannot represent, so the critics' verse-by-verse source assignments are tested as fixed hypotheses with the predictive score defined next.

3.4 Bayes factors, predictive scores, and detection power

Two hypotheses about the same counts are compared by their Bayes factor on a base-ten log scale,

$$\log_{10}\mathrm{BF}(H_1 : H_2) \;=\; \log_{10}\frac{Z(U \mid H_1)}{Z(U \mid H_2)}\,, \tag{3.6}$$

with positive values favoring H_1; at equal prior odds each unit is a factor of ten in the posterior odds. The magnitude is conditional on the word-independence model and grows with text length; the robust quantities are the side the posterior of w favors and the boundary positions that are stable across the prior settings. We call a blend decisive only when its Bayes factor exceeds the largest that any withheld single-author play of the same period receives. For a book we report the Bayes factor against a single process and the fraction of shuffled chapter orders with a Bayes factor at least as large (the permutation p of Tables 1 and 8).

Fixed assignments of units to sources are also scored by prediction of withheld units,

$$\Delta L(M_a, M_b) \;=\; \sum_{i=1}^{n}\Big[\ln P\big(U_i \mid U_{-i}^{M_a}\big) - \ln P\big(U_i \mid U_{-i}^{M_b}\big)\Big]\,, \qquad P\big(U_i \mid U'\big) \;=\; \frac{Z_\alpha(U_i + U')}{Z_\alpha(U')}\,, \tag{3.7}$$

where U_{-i}^{M} pools the other units that assignment M gives to the source of unit i, whose rates are thereby integrated out. Here ΔL is in nats (natural-log units; one $\log_{10}$ unit is 2.303 nats) and is referred to its distribution under permuted block labels.

In all three corpora the word lists, profiles, prior settings, decision rules, and published positions for comparison were fixed before any disputed text was scored, with one exception (Appendix E), and every text is reported. In each case, that ordering of events is consistent with the code and file record but is not independently provable after the fact. Not every table entry is exact: rational-arithmetic certificates cover the cells each section names, and the remaining cells run on floating-point engines checked against those cells (Appendix E). To measure detection power we insert verse by a second dramatist, a fifth of the play's verse taken from a play withheld from his profile, into a single-author play. The insertion is one contiguous stretch, one whole act, or

dispersed over the scenes, and the altered play is classified by the complete rule of Section 5.2. A play counts as long enough to test only if the insertion changes its classification at least eight times in ten over 96 replicates, for every comparison dramatist of its period and every insertion mode. Four of the 41 plays show no help at this standard. For books the inserted change matches the measured contrast between Isaiah 1–39 and 40–66 (total variation 0.192), and in synthetic books it is detected eight times in ten only from roughly 4,800 function-word tokens, the book threshold. Detections in shorter texts stand, but non-detections there are uninformative; results below hold "at measured power" in this sense.

3.5 Books with known divisions

We first run the changepoint model in one sweep over four books of known structure: Daniel, which changes language; Genesis and Micah, with divisions given in the commentaries; and, as a control, Esther, which is generally read as a single composition. We score each by chapter and by verse block at the three settings of α, on function words (Table 1) and on all lemmas (content words included), giving twelve configurations per book.

Daniel is in Hebrew through 2:4a, Aramaic from 2:4b through 7:28, and Hebrew again from chapter 8. Both boundaries (after chapters 1 and 7) carry posterior 1.0000 in every chapter-level configuration. Three processes are preferred to one at $\log_{10}$ Bayes factors of 390.1 to 541.0 on function words. The switch cannot be missed, because function words are language-specific: about 86% of Daniel's function-word types occur on only one side of it (14 of the 102 types present are shared). The Daniel result is thus the method's ceiling, reached by a language change alone: closed-class channels are robust to genre but not to language, and a cross-language changepoint says nothing about authorship.

Genesis is conventionally divided into the primeval history (chapters 1–11) and the ancestral narratives (12–50). A single process is rejected in every function-word configuration in favor of two or three, with α-dependent strength. The $\log_{10}$ Bayes factor runs from 1.2 to 53.9 and lies outside the range obtained by shuffling the chapter order at $\alpha = 1/2$ ($p < 0.001$) but inside it at $\alpha = 1/k$ ($p = 0.132$). Given three processes, one boundary falls after chapter 11 at posterior at least 0.9995 for every α and both units, while the other moves with α (Appendix A).

For Micah the commentaries offer two three-part divisions, with breaks after chapters 3 and 5 or after 2 and 5. The function-word counts support neither division, with $P(g{=}1) \geq 0.9589$ for every α and both units. However, Micah has only 982 function-word tokens. At that length a simulated language switch is detected ten times out of ten, but a contrast as strong as the Genesis division is missed thirty times out of thirty, as is a contrast half as strong. Absence of evidence at this size and channel set bounds only very large effects; it refutes nothing.

Esther stays at one process in all twelve configurations, with $P(g{=}1) \geq 0.9999$ in eleven and 0.586 in the twelfth (verse blocks, all lemmas, $\alpha = 1/2$). We draw no conclusion from the all-lemma counts (Appendix A), which respond to topic and genre as well as to a change of source. The four books are checks on the method, and we make no claim of novelty for this sweep.

4 The Federalist Papers

Twelve of the 85 Federalist papers were claimed by both Hamilton and Madison, and three more were written by the two together. We estimate a Hamilton share for every paper under the two-

book	unit	α	n	k	N	$P(g{=}1)$	$P(g{=}2)$	$P(g{=}3)$	g^*	$\log_{10}\mathrm{BF}(g^*{:}1)$	perm. p
Daniel	chapter	$1/k$	12	106	4950	0.0000	0.0000	0.9999	3	541.0	–
Daniel	chapter	1/2	12	106	4950	0.0000	0.0000	1.0000	3	486.0	< 0.005
Daniel	chapter	1	12	106	4950	0.0000	0.0000	1.0000	3	419.0	–
Daniel	block	$1/k$	23	106	4950	0.0000	0.0000	0.9995	3	501.7	–
Daniel	block	1/2	23	106	4950	0.0000	0.0000	1.0000	3	453.7	–
Daniel	block	1	23	106	4950	0.0000	0.0000	1.0000	3	390.1	–
Genesis	chapter	$1/k$	50	102	16539	0.0000	1.0000	0.0000	2	9.8	0.132
Genesis	chapter	1/2	50	102	16539	0.0000	0.0000	1.0000	3	53.9	< 0.001
Genesis	chapter	1	50	102	16539	0.0391	0.9609	0.0000	2	1.4	0.005
Genesis	block	$1/k$	78	102	16539	0.0000	1.0000	0.0000	2	8.1	–
Genesis	block	1/2	78	102	16539	0.0000	0.0000	1.0000	3	53.6	–
Genesis	block	1	78	102	16539	0.0588	0.9412	0.0000	2	1.2	–
Micah	chapter	$1/k$	7	60	982	1.0000	0.0000	0.0000	1	0.0	–
Micah	chapter	1/2	7	60	982	0.9589	0.0411	0.0000	1	0.0	1
Micah	chapter	1	7	60	982	1.0000	0.0000	0.0000	1	0.0	–
Micah	block	$1/k$	6	60	982	1.0000	0.0000	0.0000	1	0.0	–
Micah	block	1/2	6	60	982	0.9748	0.0252	0.0000	1	0.0	–
Micah	block	1	6	60	982	1.0000	0.0000	0.0000	1	0.0	–
Esther	chapter	$1/k$	10	57	2446	1.0000	0.0000	0.0000	1	0.0	–
Esther	chapter	1/2	10	57	2446	1.0000	0.0000	0.0000	1	0.0	1
Esther	chapter	1	10	57	2446	1.0000	0.0000	0.0000	1	0.0	–
Esther	block	$1/k$	12	57	2446	1.0000	0.0000	0.0000	1	0.0	–
Esther	block	1/2	12	57	2446	1.0000	0.0000	0.0000	1	0.0	–
Esther	block	1	12	57	2446	1.0000	0.0000	0.0000	1	0.0	–

Table 1: Number g of contiguous generating processes in four books of known structure, from function-word counts: Daniel's language switch is always detected, the evidence that Genesis is composite has α-dependent strength, and Micah and Esther stay at one process. Columns: unit (chapter or verse block) and Dirichlet parameter α; n units, k function words, N tokens; posterior on g (uniform prior, $g \le 6$; $P(g{\ge}4) \le 0.0005$); preferred count g^* and its $\log_{10}$ Bayes factor against $g{=}1$; chapter-order permutation p (1000 shuffles for Genesis, else 200; <: zero exceedances; 1 when $g^*{=}1$; "–": not run).

author blend of Section 3.2 and three word lists. That share is the posterior mean of w, the weight of Hamilton's profile in the blend, both profiles being fixed at their pooled values. It is a share of the paper's word-list tokens, within 0.0050 of the corresponding share of running text, and not a calibrated fraction of authorship.

4.1 Hamilton's list and Madison's

The Federalist essays ran in New York newspapers in 1787–88 over the signature Publius. Hamilton left a handwritten allocation the day before his duel with Burr; Madison's list appeared essay by essay in the Gideon edition of 1818 (Mosteller and Wallace, 1964). The two lists disagree on twelve essays; Nos. 18–20 are acknowledged joint work, and five are Jay's. The historian Douglass Adair (1944) decided for Madison on documentary evidence.

Mosteller and Wallace (1963, 1964) found the two men "practically twins" in sentence length (means of 34.55 and 34.59 words) and separated them instead by the rates of function words. *Upon*,

word list	selection	k	correct	H	M	mean margin	worst paper
M&W 70 (continuity)	a priori	70	64/65	50/51	14/14	+15.235	−1.426
closed class (cc318)	a priori	318	64/65	51/51	13/14	+22.502	−10.597
top-70	refit	70	63/65	50/51	13/14	+17.649	−4.329
top-300	refit	300	64/65	51/51	13/14	+32.671	−18.170
top-1000	refit	1000	62/65	51/51	11/14	+46.744	−29.890
full vocabulary	refit	7805–7912	53/65	51/51	2/14	+86.073	−97.703
M&W 30 (their final list)	M&W	30	65/65	51/51	14/14	+15.520	+3.047

Table 2: Seven word lists compared by leave-one-paper-out prediction on the 65 papers of known authorship: three lists give 64 of 65 papers the correct sign, longer data-derived lists misassign Madison papers, and Mosteller and Wallace's final thirty-word list, selected on these papers, makes no error. "Correct": left-out papers whose pure-model $\log_{10}$ Bayes factor has the true author's sign, by Hamilton (H) and Madison (M) papers; margins: that factor toward the true author (mean, worst paper); k: list length; selection: fixed a priori, refit per left-out paper, or M&W's own.

for example, occurs at three per thousand words in Hamilton and at one sixth per thousand in Madison. Mosteller and Wallace gave all twelve disputed essays to Madison, at posterior odds that they themselves reduced to allow for model uncertainty. For the joint papers, where in their words "the issue is the extent of each man's contribution," they reported whole-paper log odds, judged Nos. 18 and 19 mainly Madison's work, and found No. 20 a puzzle.

Most later studies assign each whole paper to a single author. Collaboration on the disputed papers has been argued from rhetoric and stylometry (Collins et al., 2004; Rudman, 2012; Ebrahimpour et al., 2013), and the joint papers stand out in vocabulary richness (Holmes and Forsyth, 1995), but these studies, like recent classifiers built on language models (Jeong and Ročková, 2025), do not estimate the size of a second contribution. Fully Bayesian analyses in the Mosteller–Wallace line likewise treat each text as one author's (Airoldi et al., 2006; Gill and Swartz, 2011; Puig et al., 2016; Bozza et al., 2025; Hjort, 2007).

4.2 Word lists on the undisputed papers

The 65 papers of known authorship allow any word list to be graded by how well it predicts the author of a paper left out of the profiles (Table 2). The list inherited from Mosteller and Wallace ("continuity" in the tables) is the seventy function words they screened as candidate markers (Mosteller and Wallace, 1964; Fung, 2003); their final list, a shorter refinement, we reconstruct as thirty words with variant forms merged. The closed-class list (cc318) is a 318-word English stop-word list taken unmodified from a software library.[1] Three lists take the 70, 300, and 1000 most frequent words of the training papers; another is the full vocabulary.

We left each paper of known authorship out in turn (the joint, disputed, and Jay papers excluded), rebuilt the profiles and the frequency-ranked lists from the rest, and computed for the left-out paper the $\log_{10}$ Bayes factor between the two pure-author models. We ranked the lists first by the number of correct signs and then by the mean Bayes factor toward the true author, a rule fixed in advance.

The inherited list, the closed-class list, and top-300 each give 64 of 65 papers the correct sign.

[1]The English stop-word list of the scikit-learn library, version 1.9.0.

paper	word list	N	$E[w]$	$q_{0.05}$	$q_{0.5}$	$q_{0.95}$	$\log_{10}$ BF(blend : best pure)
18	M&W 70 (continuity)	1003	0.0368	0.0022	0.0277	0.1022	−1.361
	closed class (cc318)	1192	0.0665	0.0060	0.0570	0.1600	−0.911
	top-300	1301	0.0709	0.0064	0.0613	0.1689	−0.880
19	M&W 70 (continuity)	962	0.0579	0.0034	0.0437	0.1610	−1.171
	closed class (cc318)	1130	0.1604	0.0275	0.1535	0.3172	−0.169
	top-300	1272	0.2297	0.0830	0.2280	0.3815	+0.755
20	M&W 70 (continuity)	697	0.1183	0.0107	0.1006	0.2872	−0.655
	closed class (cc318)	831	0.2078	0.0499	0.2013	0.3872	+0.243
	top-300	967	0.2579	0.0973	0.2549	0.4282	+0.984

Table 3: Posterior of the Hamilton weight w in the three joint papers under each evaluation list: every posterior is Madison-dominant, with means comparable to or below the share that a paper Madison wrote alone receives when left out of the profiles and analyzed the same way (median 0.13, 0.19, 0.26; maximum 0.43, 0.60, 0.73). N: tokens of the paper in the list; $E[w]$: posterior mean (the Hamilton share); q: posterior quantiles; last column: $\log_{10}$ Bayes factor of the blend against the better pure-author model. Continuity rows use the inherited list.

The inherited list misassigns Hamilton's No. 6 narrowly, the closed-class list Madison's No. 38 firmly, and top-300 Madison's No. 10, where about a quarter of the wrong-direction evidence comes from content words (share 0.266). Under the full vocabulary the number correct falls to 53 of 65; every error is a Madison paper assigned to Hamilton, the worst being No. 10 again, at posterior odds of about 10^{98}. There content words led by *faction* carry about half the wrong-direction evidence (share 0.507), so a list that long appears to track subject as much as author. The thirty-word list makes no error, but its words were chosen for discriminating power on these same papers, so its accuracy here is optimistic. With that list set aside, the rule ranks top-300 first, the closed-class list second, and the inherited list third; we keep all three and call the first two the winning lists.

4.3 The joint papers and No. 55

The three joint papers are Madison-dominant under every list (Table 3 and Fig. 1). The posterior mean of the Hamilton weight w spans 0.037 to 0.071, 0.058 to 0.230, and 0.118 to 0.258 for Nos. 18, 19, and 20 across the three lists, the larger under the two winning lists. Left out of the profiles in turn and analyzed the same way, Madison's undisputed papers receive Hamilton shares with median 0.13, 0.19, and 0.26 under the three lists. Thus the Hamilton share of the joint papers is comparable to, or below, that of a typical paper by Madison alone; no Hamilton contribution above that level is resolved in them. Under the hierarchical model, which admits the paper-to-paper rate variation Mosteller and Wallace found, both sets of numbers fall (the largest joint-paper mean to 0.0941) and the comparison is unchanged (Table 13).

No. 20, which puzzled Mosteller and Wallace, has the largest Hamilton share of the three under every list; its blend factor against the better pure author changes sign with the list, a weak preference either way, as does No. 19's. Split in halves or thirds, all three joint papers show internal variation consistent with the overdispersion Mosteller and Wallace documented (Appendix D).

Of the twelve disputed papers (Table 4), only No. 55 has a Hamilton share above one half under any list. Under the closed-class list and top-300 its shares are 0.3198 and 0.3879, and the

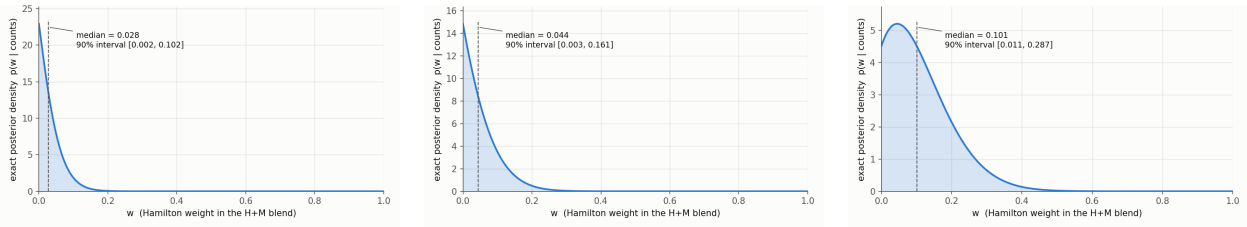

Figure 1: Posterior density of the Hamilton weight w for the joint papers Nos. 18, 19, and 20 (left to right) under the inherited 70-word list. The horizontal axis is w, the weight of Hamilton's profile p_A against Madison's p_B in the blend $w\,p_A + (1-w)\,p_B$, and the vertical axis is the posterior density; the dashed line marks the posterior median. All three densities peak at small w, No. 20's with the largest median and spread; under the closed-class list and top-300 the densities for Nos. 19 and 20 move to larger w (Table 3).

pure-author comparison favors Madison ($\log_{10}\mathrm{BF}(H{:}M) = -9.313$ and -6.764), which gives a Madison majority under the two winning lists, with a Hamilton share near a third. However, under the inherited list the share is 0.5146 and the pure-author comparison leans to Hamilton (+0.605), so the Madison-majority reading is the winning-sets reading, not the continuity column's. Under Mosteller and Wallace's own thirty-word list the comparison is on Madison's side again (-1.14, share 0.41), and re-estimated profiles spread the inherited list's lean across zero (Appendix D). Puig et al. (2016), with closed-form Dirichlet–multinomial evidences for Hamilton, Madison, and an unnamed third author, likewise found No. 55 alone among the disputed papers drawn toward Hamilton under a list chosen from the papers, and on Madison's side under Mosteller and Wallace's list.

For No. 55 the blend is preferred under all three lists, by factors of $10^{2.0}$ to $10^{4.0}$. Yet that factor is not by itself evidence of mixture: 3, 6, and 9 of the 65 left-out single-author papers reach it under the three lists. The blend's extra freedom also absorbs rate variation between papers that the word-independence model omits, and a share near a third may reflect either a Hamilton contribution or that variation; these data do not separate the two. Under the hierarchical model, which represents that variation, the preference disappears and the Hamilton share under the two winning lists falls to 0.0749 and 0.1322 (Table 13). That model, however, detects an even mixture of two left-out papers at 10^2 in only 19 of 30 trials (3 of 10 under the inherited list), against 29 of 30 under word independence (Appendix D). Neither the preference nor its disappearance therefore bears on whether No. 55 is joint work.

Weighted by token counts over the 80 Hamilton and Madison papers (Jay's five excluded), the shares give Hamilton about three-fifths of the Federalist under either model: under word independence $F = \sum_i N_i\,E[w_i] / \sum_i N_i$ is 0.5895, 0.6145, and 0.6219 by list. The evidence integral is that of Lin et al. (2009); weight posteriors for mixtures of known components have been computed in other applications (Meshcheryakov, 2026; Kim, 2026). As far as two independent literature searches with distinct query families could find, no blend hypothesis with a mixing weight had ever entered any Federalist hypothesis space, Bayesian or otherwise. Nor did they find a mixing-weight posterior for any composite text in stylometry; those computed here appear to be the first.

Except where the hierarchical model is named, this section uses the word-independence model, the idealization Mosteller and Wallace rejected because the counts overdisperse. Refitted under the hierarchical model, every disputed and joint paper is on Madison's side, most at roughly half

	M&W 70 (continuity)				closed class (cc318)				top-300			
paper	$E[w]$	$q_{0.05}$	$q_{0.5}$	$q_{0.95}$	$E[w]$	$q_{0.05}$	$q_{0.5}$	$q_{0.95}$	$E[w]$	$q_{0.05}$	$q_{0.5}$	$q_{0.95}$
49	0.3270	0.0971	0.3229	0.5691	0.3461	0.1595	0.3450	0.5357	0.2644	0.1328	0.2636	0.3987
50	0.2200	0.0356	0.2057	0.4540	0.1190	0.0126	0.1054	0.2730	0.1731	0.0345	0.1661	0.3357
51	0.0911	0.0065	0.0740	0.2347	0.0565	0.0036	0.0439	0.1526	0.0328	0.0020	0.0249	0.0908
52	0.0954	0.0064	0.0763	0.2495	0.1087	0.0101	0.0958	0.2525	0.1156	0.0134	0.1063	0.2510
53	0.0973	0.0068	0.0789	0.2509	0.1552	0.0251	0.1480	0.3107	0.1980	0.0636	0.1962	0.3378
54	0.1973	0.0432	0.1879	0.3832	0.2187	0.0672	0.2138	0.3865	0.2429	0.1033	0.2407	0.3896
55	**0.5146**	**0.2811**	**0.5136**	**0.7517**	**0.3198**	**0.1667**	**0.3191**	**0.4750**	**0.3879**	**0.2460**	**0.3876**	**0.5308**
56	0.1663	0.0167	0.1480	0.3805	0.0444	0.0026	0.0334	0.1236	0.2322	0.0886	0.2299	0.3831
57	0.0857	0.0061	0.0696	0.2207	0.1791	0.0373	0.1740	0.3383	0.3071	0.1671	0.3065	0.4492
58	0.1657	0.0189	0.1514	0.3632	0.2098	0.0506	0.2057	0.3821	0.1847	0.0495	0.1819	0.3288
62	0.1196	0.0095	0.1014	0.2924	0.1935	0.0389	0.1883	0.3660	0.2859	0.1385	0.2857	0.4341
63	0.1410	0.0200	0.1323	0.2930	0.1664	0.0437	0.1634	0.2990	0.3492	0.2344	0.3487	0.4655

Table 4: Posterior of the Hamilton weight w in the twelve disputed papers under the three evaluation lists, as its mean $E[w]$ and its 5th, 50th, and 95th percentiles: No. 55 (bold) is the only disputed paper whose mean exceeds one half under any list, and only under the inherited list (continuity); under the closed-class list the largest mean is No. 49's (0.3461). Every other disputed paper is Madison-dominant under all three lists.

the Bayes factor, and the left-out single-author fits whose blend factor reaches 10^2 fall from 24 to 6 of 195 (65 papers under three lists; Appendix D). Ninety-percent intervals for w on left-out single-author papers, meant to miss the truth one time in ten, miss it one time in three under word independence and about one in four under the hierarchical model. In consequence the Bayes-factor magnitudes are conditional on the count model, and of the blend-weight posterior only the side it favors, not its location or width, is robust to the profile, prior, word-list, and dispersion checks we ran. The results are comparisons of blend and pure-author models, and we make no attribution claim beyond the exact evidence numbers, under any channel set.

5 The Shakespeare canon

The open questions of authorship in the Shakespeare canon concern collaboration: which plays had a second hand, and where. We score all forty-one plays against the profiles of eight other dramatists under one rule, then take up the hardest case, the three parts of *Henry VI*.

5.1 The collaboration question

Doubt about single authorship is as old as Shakespeare editing. Malone (1787), the canon's first modern editor, argued that the three *Henry VI* plays were not originally Shakespeare's. Hickson (1847) divided *The Two Noble Kinsmen* between Shakespeare and Fletcher, and Spedding (1850) split *Henry VIII* scene by scene on counts of eleven-syllable lines. The New Shakspere Society of 1873 set out to make such counting scientific, and its successors reassigned so much of the canon that Chambers's lecture *The Disintegration of Shakespeare* made the enterprise disreputable for half a century (Chambers, 1924). Modern attribution scholarship reversed that judgment: Hoy's division of *Henry VIII* and *The Two Noble Kinsmen* (Hoy, 1956), Vickers's synthesis of five collaborations (Vickers, 2002), Jackson's case for Wilkins's hand in *Pericles* (Jackson, 2003),

and Craig and Kinney's computational studies (Craig and Kinney, 2009). In 2016 the New Oxford Shakespeare printed Marlowe's name beside Shakespeare's on the *Henry VI* title pages and gave co-author credits on seventeen of its forty-four works (Taylor and Egan, 2017). Whether Shakespeare wrote Shakespeare is a separate and settled question (Edmondson and Wells, 2013).

5.2 All forty-one plays

For each play we test a blend of Shakespeare's profile for the period with one of eight other dramatists' profiles against the better profile alone. Every number in this section is a model-comparison evidence object over frozen profiles, never an authorship attribution. Undisputed and disputed plays are scored alike, making the survey a census of the canon.

Five of the eight comparison dramatists have at least two securely solo plays and twenty thousand function-word tokens: Fletcher (five plays), Middleton (six), Peele (four), Marlowe (five), and Greene (four). Three further profiles rest on one play each and support no conclusion: Wilkins, Nashe (4,576 tokens; he wrote mostly prose), and Kyd, whose one secure play is *The Spanish Tragedy*.[2] Shakespeare's profile and the comparison dramatists are matched by period to each play, with 27 plays dated before 1603 and 14 after; all profiles use one list of 365 function words and are held fixed. The eight comparison profiles form the pool: any hand outside these eight frozen profiles is undetectable by construction, and every null below is a null against the pool, never against anyone.

For each play and comparison profile (177 pairs) we compute the blend gain M, the $\log_{10}$ Bayes factor of the best blend with Shakespeare's profile over the best single profile. We take the smallest value obtained as the plays of either profile are left out one at a time. The ceiling C_{reg} is the largest M any securely solo play of the same period reaches when held out and scored the same way. We class a play as having had help when M exceeds C_{reg} by a fixed margin and at least one act is decisive, favoring the comparison profile by itself beyond a threshold calibrated on the acts and scenes of solo plays (Appendix B). We class a play as no help only where it meets the power requirement of Section 3.4, that a planted fifth by any comparison dramatist would change the classification, as it would for five of the thirteen plays tested. Plays in neither class are unresolved ("cannot resolve at power"), and for these, absence of signal below that power is never read as absence of help.

Under this rule six plays read had-help, four show no help at measured power, and thirty-one cannot be resolved (Table 5). The six are *3 Henry VI*, *Edward III*, *Henry VIII*, *The Two Noble Kinsmen*, *Titus Andronicus*, and *2 Henry VI*. The four are *King John*, *A Midsummer Night's Dream*, *Richard III*, and *The Tempest*. For these no profile clears its ceiling, no act is decisive, and a planted fifth by any comparison dramatist moves the play out of the "none" class at least 84 percent of the time, mostly to "unresolved" (Appendix B). With a tenth of the verse planted instead of a fifth, only *A Midsummer Night's Dream* still meets the requirement against every comparison dramatist, and *King John* and *Richard III* only against Peele; help confined to prose is outside the test (Appendix E).

Of the thirty-one unresolved plays, eight fail the power requirement. In seven Elizabethan comedies and tragedies, *Hamlet* among them, a planted fifth of Greene's or Marlowe's verse moves the play out of the "none" class as rarely as 0.09 to 0.52 of the time. In *Coriolanus* a fifth of Middleton's does so 0.12 of the time. In sixteen we planted no text, because a simpler calculation

[2] *Cornelia* (a translation), *Soliman and Perseda*, and the *First Part of Hieronimo* (both disputed) were excluded.

play	profile	N	M	C_{reg}	power	result	position in the literature
3 Henry VI	Marlowe	9496	53.47	7.58	–	**had help: act 1**	revision or collaboration debated (Malone, 1787; Craig and Kinney, 2009; Taylor and Egan, 2017)
Edward III	Greene	7344	46.59	1.22	–	had help: acts 3, 5	Shakespeare in part; Kyd proposed
Henry VIII	**Middleton**	6710	62.55	19.58	–	had help: acts 4, 5, prologue	with Fletcher (Spedding, 1850; Hoy, 1956; Plecháč, 2021)
Two Noble Kinsmen	Middleton	6272	56.38	19.58	–	had help: acts 2, 4, epilogue	with Fletcher (Hickson, 1847; Hoy, 1956)
Titus Andronicus	Peele	4273	27.98	1.92	–	had help: **act 3**	with Peele (act 1; 2.1, 4.1) (Vickers, 2002)
2 Henry VI	Marlowe	8266	27.88	7.58	–	**had help: act 1**	revision or collaboration debated (Craig and Kinney, 2009; Taylor and Egan, 2017)
King John	–	4104	−0.10	1.92	0.84	none at this length	solo
Midsummer Night's Dream	–	4956	1.17	1.22	0.93	none at this length	solo
Richard III	–	11404	1.68	7.58	0.86	none at this length	solo
Tempest	–	4553	0.64	2.36	0.90	none at this length	solo
All's Well That Ends Well	–	3637	4.27	19.58	–	power unmeasured	Middleton share disputed
Antony and Cleopatra	–	5573	3.11	19.58	–	power unmeasured	solo
Arden of Faversham†	–	6681	13.75	1.22	–	gain not localized	Shakespeare in part (Craig and Kinney, 2009); Kyd proposed
As You Like It	–	3770	−1.19	1.22	0.29	power too low	solo
Comedy of Errors	–	5084	−1.88	1.22	0.09	power too low	solo
Coriolanus	–	6054	0.83	19.58	0.12	power too low	solo
Cymbeline	–	7121	2.89	19.58	–	power unmeasured	solo
Hamlet	–	9227	−1.80	7.58	0.15	power too low	solo
1 Henry IV	–	6002	1.36	7.58	–	power unmeasured	solo
2 Henry IV	–	5279	0.21	7.58	–	power unmeasured	solo
Henry V†	–	6547	6.09	7.58	–	gain not localized	solo
1 Henry VI†	–	8267	26.65	7.58	–	gain not localized	Nashe (act 1), others; Kyd (Vickers, 2011)
Julius Caesar	–	7627	−1.63	7.58	–	power unmeasured	solo
King Lear	–	5573	−1.09	19.58	–	power unmeasured	solo
Love's Labour's Lost	–	5583	−1.80	1.22	0.20	power too low	solo
Macbeth	–	4605	−1.32	19.58	–	power unmeasured	adapted by Middleton
Measure for Measure	–	4214	−0.29	19.58	–	power unmeasured	adapted by Middleton
Merchant of Venice	–	6739	−1.84	1.22	0.22	power too low	solo
Merry Wives of Windsor	–	992	3.78	7.58	–	power unmeasured	solo
More, Hand D†	–	99	2.03	1.92	–	gain not localized	Shakespeare's addition
Much Ado About Nothing	–	1441	−1.14	1.92	–	power unmeasured	solo
Othello	–	6605	10.28	19.58	–	power unmeasured	solo
Pericles	–	4420	5.57	19.58	–	power unmeasured	with Wilkins (acts 1, 2) (Vickers, 2002; Jackson, 2003)
Richard II†	–	9517	7.58	6.09	0.96	gain not localized	solo
Romeo and Juliet†	–	5050	1.92	0.37	–	gain not localized	solo
Taming of the Shrew	–	6740	−0.30	1.22	0.52	power too low	solo
Timon of Athens	–	4095	0.16	19.58	–	power unmeasured	with Middleton (act 3) (Vickers, 2002)
Troilus and Cressida	–	7306	0.05	7.58	–	power unmeasured	solo
Twelfth Night	–	3041	−1.64	1.22	–	power unmeasured	solo
Two Gentlemen of Verona	–	5396	−1.12	1.22	0.32	power too low	solo
Winter's Tale†	–	5577	19.58	3.11	–	gain not localized	solo

Table 5: All forty-one plays against eight comparison dramatists, by block: second hand detected; none at this length; unresolved. N: function-word tokens. M: $\log_{10}$ Bayes factor of the best blend of Shakespeare's profile with another dramatist's (named in the first block) over the best single profile; result: the acts decisive by themselves. C_{reg}: largest M of a held-out solo play of the period. Power: rate at which a planted fifth by the least detectable dramatist moves the play from "none" under the complete rule ("none" requires at least 0.8; a detection does not; –, not measured). †: a gain above some profile's C_{reg} localized to no single pair. Last column: position in the attribution literature (after Taylor and Egan (2017) unless cited); bold marks a departure.

with a simulated fifth falls short or is marginal (Appendix B). Nine of these are late plays, which face the Middleton-side ceiling of 19.58, and the rest are short or face the Marlowe ceiling of 7.58. In the other seven, marked †, the blend gain exceeds some ceiling but matches no single pair of profiles. Cannot resolve at power is a statement about the instrument's measured resolving power at that play's N and ceiling, never about the play's history and never a doubt cast on anyone's claim. Two profile-free analyses detect almost nothing (Appendix B).

The classifications of Table 5 assume the word-independence model. Under the hierarchical model the blend gains of the six fall at least threefold, most ceilings fall too, and the same rule returns four plays with a second hand, *King John* alone with none, and thirty-six unresolved (Table 11). *Henry VIII* and *The Two Noble Kinsmen* keep decisive Middleton-side acts, and *Edward III* and *Titus Andronicus* still qualify, though now through the Marlowe profile (acts 3 and 1). Parts Two and Three of *Henry VI* keep gains of 8.77 and 16.16 against a Marlowe ceiling and margin of 2.83 and 1.59, but no single act stays Marlowe-side, so both read "gain not localized." The other three "none" plays become unresolved through that model's lower ceilings and power, and one of the fifteen solo plays behind the ceilings, *Henry V*, falls on Marlowe's side, as none does under word independence.

The last column of Table 5 maps each play's position in the attribution literature. The map changes no verdict; the comparison is descriptive, and no cell of it is an attribution. For eight of the ten resolved plays the result, help or none, agrees with the published position, and for two it is stronger. No play that the literature divides is among the four, and none that it regards as solo is among the six. The decisive acts of *The Two Noble Kinsmen* and *Henry VIII* fall where Spedding (1850) and Hoy (1956) placed the second hand, and those of *Edward III* lie outside the Countess scenes, its most securely Shakespearean part. For *Titus Andronicus* the comparison profile, Peele, agrees with Vickers (2002), but the decisive act is act 3, whereas the published division gives Peele act 1 with scenes 2.1 and 4.1. The Marlowe profile also clears its ceiling for *Titus* (21.99 against 7.58) and is decisive in act 1, and under the hierarchical model that act alone carries the classification.

In the scholarship on *Henry VIII* the second hand is Fletcher's, from Spedding (1850) and Hoy (1956) to the word and rhythm analysis of Plecháč (2021). The Fletcher profile barely clears its ceiling (3.34 against 2.36) and is decisive only in the prologue, while the Middleton profile is decisive at $M = 62.55$: on function words the second hand is closer to the Middleton profile than to Fletcher's. Under the hierarchical model the Fletcher blend gain turns negative (-2.18) while the Middleton gain (9.82 against a ceiling of 4.64) and its decisive act 4 and prologue remain. This need not mean a different collaborator, since two dramatists writing in the same late idiom may each fit the other's profile, and the late-Shakespeare concentration rests on four plays. In *The Two Noble Kinsmen* the Fletcher profile is decisive in its own right (18.20 against 2.36, act 2). The two results stronger than the published positions concern the later *Henry VI* plays, entered in Table 5 (Malone, 1787; Craig and Kinney, 2009; Taylor and Egan, 2017) as mainly Shakespeare's with revision or collaboration debated. In both the blend with Marlowe is decisive in act 1 under word independence; in Section 5.3 we rebuild the profiles under a stricter standard of sole authorship and ask on which dramatist's side each act falls.

Eight plays that the literature divides are unresolved. In *1 Henry VI* the blend gain exceeds every ceiling but localizes to no single pair. Nashe's proposed hand in its act 1 (Taylor and Egan, 2017), Wilkins's in *Pericles* (Jackson, 2003), and Kyd's in *Arden of Faversham* and *Edward III* cannot be tested with one-play profiles, though the Kyd profile does clear its ceiling for both. The

play	N	$\log_{10}$BF(S:M) [band]	$\log_{10}$BF(S:K)	blend	w [q5, q95]
1 Henry VI	4477	+47.2 [+35.8, +51.5]	+54.5	+9.9	0.70 [0.62, 0.77]
2 Henry VI	4476	+35.3 [+22.8, +37.7]	+55.1	+12.7	0.64 [0.57, 0.72]
3 Henry VI	5045	−1.2 [−10.0, +2.3]*	+36.8	+30.9	0.48 [0.41, 0.55]

Table 6: The three parts of *Henry VI*: in each a Shakespeare–Marlowe blend is favored over the best single profile and Kyd is never preferred, while the Shakespeare weight w falls across the trilogy. Columns: verse function-word tokens N (142-word list); $\log_{10}$ Bayes factor, Shakespeare against Marlowe profile (positive favors Shakespeare), with its band, the range over profiles omitting one reference play other than *Edward II*; the same against Kyd; $\log_{10}$ Bayes factor, blend against best single profile; posterior mean and 5th/95th percentiles of w. *Band includes zero; see text.

Middleton adaptations proposed for *Measure for Measure* and *Macbeth* are smaller than a scene. Twenty-three plays that the literature treats as solo are unresolved; the census cannot confirm them, which is different from doubting them.

Throughout, "a second hand" means a better fit by two profiles than by one, and a reviser would not be distinguished from a collaborator. The words counted were chosen to be insensitive to differences of edition. The magnitudes of M, and the Elizabethan decisive acts for the profiles named in Table 5, hold only under the word-independence model; the four detections other than the *Henry VI* plays, and *Henry VIII*'s preference for Middleton over Fletcher, hold under both.

5.3 The three parts of Henry VI

The authorship of the three parts of *Henry VI* has been in doubt since Malone (1787). The New Oxford Shakespeare gives act 1 of Part One to Nashe, with Marlowe queried elsewhere in that play, and places Marlowe's hand in Parts Two and Three (Taylor and Egan, 2017). Craig and Kinney (2009) associate the Jack Cade scenes of Part Two, and stretches of Part Three, with Marlowe. Vickers (2011) has claimed Part One for Kyd, and Segarra et al. (2016) classify the trilogy by word-adjacency networks.

We compare the verse of each play, act, and scene with profiles of Shakespeare, Marlowe, and Kyd. Shakespeare's profile comes from six history plays of undisputed sole authorship and Marlowe's from the two parts of *Tamburlaine* and *Edward II* (a sensitivity variant adds *Dido* and *The Jew of Malta*). Kyd's comes from *The Spanish Tragedy* alone, whose 1592 quarto marks acts but no scenes, which precludes a scene-level Kyd comparison. The word list (142 function words) was chosen under a rule fixed in advance, by leave-one-out testing on the nine reference plays, each withheld in turn from the profile of its author; the list classifies eight of the nine correctly. The exception is *Edward II*, Marlowe's one undisputed history play in dialogue form, misclassified with a Bayes factor of 10^{275}. No reference set we tried classifies every play correctly, as we had required (Appendix B). The results are therefore reported under a disclosed override of a pre-registered validation bar: every Marlowe-side number in this section is conditional on the coverage of one profile play, *Edward II*, and the Kyd comparison carries the diagnostic grade of a one-play profile.

In all three parts the evidence favors a Shakespeare–Marlowe blend over the best single profile (Table 6), by factors of $10^{9.9}$, $10^{12.7}$, and $10^{30.9}$, and the posterior Shakespeare weights are 0.70, 0.64, and 0.48. Withheld from their own profiles and compared the same way, the nine reference plays receive blend factors from $10^{-1.6}$ to $10^{38.8}$. The largest is that of *Edward II*, a held-out pure

Marlowe play, with posterior weight $w = 0.49$. *The Spanish Tragedy*, in neither profile, receives $10^{19.7}$; *Dido* and *The Jew of Malta*, whose Marlowe authorship is itself contested, receive $10^{10.9}$ and $10^{8.8}$. The blend preference alone therefore does not separate collaboration from unmodeled style: three, two, and one of these twelve reference values exceed the factors for Parts One, Two, and Three, respectively. Under the hierarchical model the factors become $10^{4.0}$, $10^{4.9}$, and $10^{11.8}$, with w of 0.70, 0.64, and 0.46, and *Edward II*'s falls to $10^{4.5}$ while its Shakespeare-against-Marlowe factor turns Marlowe-side (−7.3). One, none, and none of the twelve reference values then exceed those of the three parts, so that only Part Three's preference stands clear of what solo plays produce, at a Marlowe concentration estimated from two or three plays. The Kyd profile, built from one play, is never preferred, which counts against Vickers's proposal for Part One without settling the question.

Part Three's $\log_{10}$ Bayes factor for Shakespeare against Marlowe is −1.2, and its robustness band (the range over profiles that omit one reference play other than *Edward II*) is $[-10.0, +2.3]$. This play-level number is a straddle (its band includes zero); under each of three alternative word lists the same comparison favors Shakespeare (+36.1, +42.2, +7.2), and without *Edward II* in the Marlowe profile it gives +49.5. We therefore summarize Part Three as a blend, $w = 0.48$ [0.41, 0.55]. Its Marlowe-side evidence lies in acts 1 and 4 (−7.2 and −7.8), and these too favor Shakespeare once *Edward II* is removed (+6.0, +0.8).

In Fig. 2 we class an act or scene as Shakespeare-side or Marlowe-side when its whole robustness band lies beyond nearly all values obtained for reference scenes of the other dramatist at similar length, and uncertain otherwise. Act 1 of Part Two, where the blend with Marlowe was decisive in Section 5.2, falls on neither side here. Part Two's one Marlowe-side act is act 4 (Craig and Kinney's Cade scenes), and its band, $[-5.0, -3.3]$, clears the threshold by only 0.11. Without *Edward II* the act favors Shakespeare (+2.4), and the result rests on the act's verse alone (727 function-word tokens), though much of the Cade material is prose. We claim no more of the agreement with Craig and Kinney than these conditions support.

Of the trilogy's 77 scenes, 73 have fewer than the $N = 400$ function-word tokens at which reference scenes of the two dramatists separate cleanly. At these lengths most scenes of a securely solo play would read uncertain here too, and an uncertain class probably reflects the scene's length; the usable evidence is in the acts and plays.

The Marlowe-side evidence falls where the New Oxford editors (Taylor and Egan, 2017) and Craig and Kinney (2009) place Marlowe, in Part Three and the Cade act. Yet the *Edward II* result shows that under word independence a blend preference as strong as any part's may arise without a second hand. In Part One the evidence favors Shakespeare over Marlowe and Kyd, but without a Nashe profile in the comparison, Part One's Shakespeare-side reads cannot refute the Nashe attribution of its first act.

6 The Bible

In the Protestant canon the Bible comprises sixty-six books, in Hebrew, Aramaic, and Greek. We take the Hebrew and Aramaic text from the Leningrad Codex of 1008 CE and the Greek from the SBL Greek New Testament. We ask where the Book of Isaiah changes source in two manuscripts a millennium apart (Section 6.2); whether the Pentateuchal sources separate on function words (Section 6.3); and how many generating processes each book contains, and how many distinguishable profiles all sixty-six together (Section 6.4). We are counting processes, never authors, since a

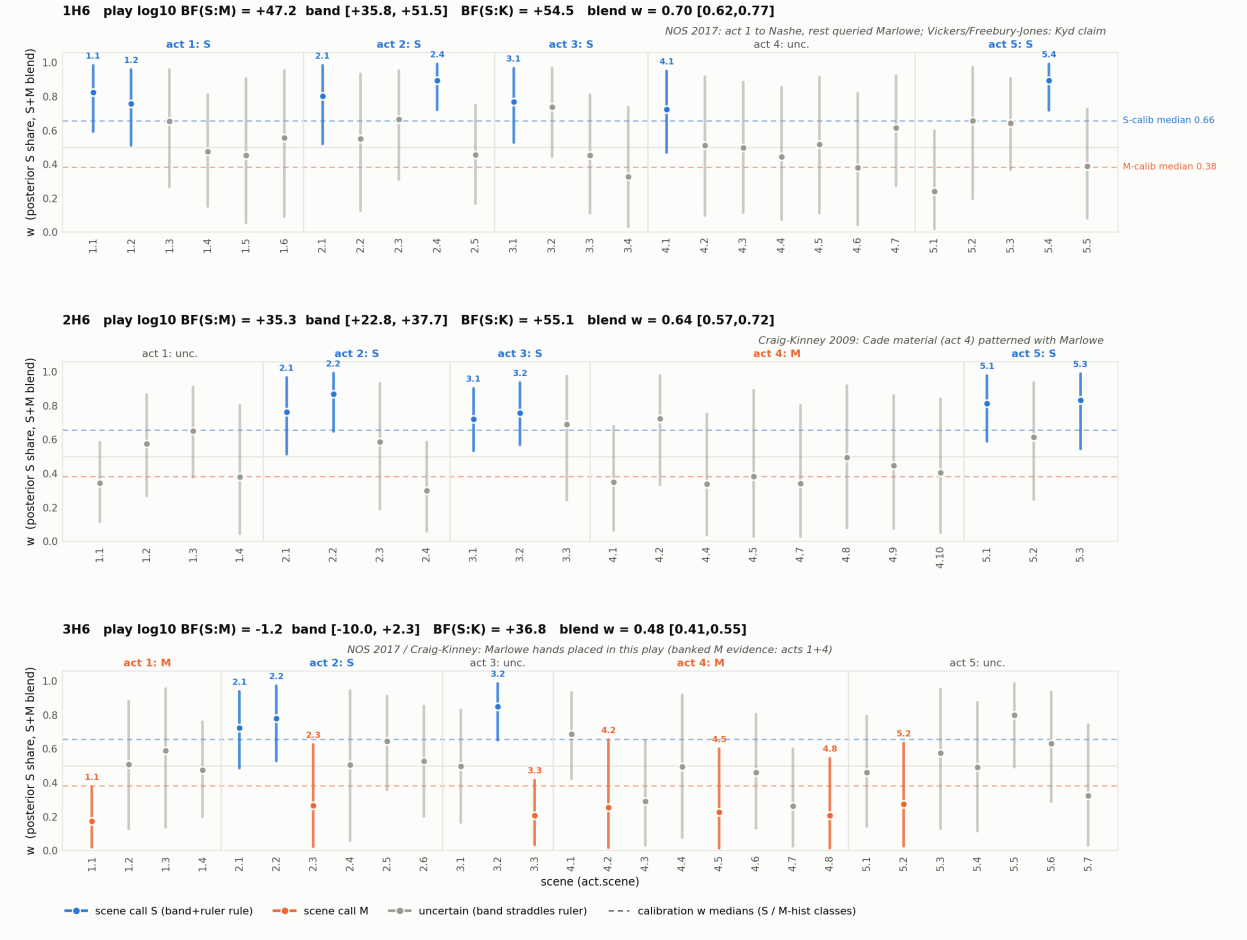

Figure 2: Posterior Shakespeare weight w (dot, mean; bar, 5th to 95th percentile) in the Shakespeare–Marlowe blend for every scene of the three parts of *Henry VI*, by act: Marlowe-side classes occur only in Part Three and in act 4 of Part Two. Blue marks Shakespeare-side scenes and orange Marlowe-side (whole band beyond nearly all reference-scene values of the other dramatist at that length); gray, uncertain. Act classes appear above each act and play-level values (Table 6) above each panel; dashed lines mark the two dramatists' reference-scene medians; italics give published placements.

change of genre or language within a book may alter its function-word rates as much as a second writer would.

6.1 Source criticism since Astruc

Jean Astruc (1753), writing anonymously, suggested that Genesis alternates between two names of God and that Moses had spliced together older memoirs, which Astruc printed in two parallel columns. De Wette (1805) set Deuteronomy apart as a later document, and Hupfeld (1853) found two sources in the Elohim material of Genesis. Wellhausen, who wrote that he had accepted Graf's late dating of the priestly law "almost without knowing his reasons," gave the four-source hypothesis its classical form in 1878 (Wellhausen, 1885). Carpenter and Harford-Battersby (1900) later tabulated the assignment of each verse to a source. The documentary hypothesis itself is contested: Schmid (2021) distinguishes documentary, neo-documentary, and supplementary positions in current Pentateuchal scholarship.

Isaiah was divided in the eighteenth century as well: Döderlein (in the 1770s and 1780s) and Eichhorn (in 1783) assigned chapters 40 onward to a later, exilic author (Döderlein, 1789). Duhm (1892) cut the book in three: chapters 1–39, 40–55 (Deutero-Isaiah), and 56–66 (Trito-Isaiah). This division was the consensus for most of the twentieth century. The Great Isaiah Scroll (found at Qumran in 1947 and containing all sixty-six chapters) was copied around 125 BCE, when the book was already regarded as one composition (Ulrich and Flint, 2010).

Among the epistles, Semler (1776) questioned the unity of Second Corinthians. Baur (1845) reduced the Pauline corpus to the four letters he found beyond suspicion (Galatians, the two Corinthian letters, and Romans), which he called the Hauptbriefe. For the gospels, Holtzmann (1863) won general acceptance for Markan priority combined with a second source consisting of sayings, a scheme that Streeter (1924) elaborated into a four-document hypothesis.

Statistical tests of the source critics' divisions go back at least to Wickmann (1968), whose method was a Bayes decision between same and different authorship for a fixed pair of texts. With computer-assisted multivariate tests, Radday (1970, 1973) found Isaiah 40–66 statistically distinct from 1–39, and Radday and Shore (1985) extended the approach to Genesis. In the 2010s, composite books were decomposed with classifiers and clustering (Koppel et al., 2011; Akiva and Koppel, 2013; Dershowitz et al., 2015).

6.2 Isaiah on two manuscripts

Duhm (1892) divided Isaiah among three hands; rather than fix the number of sources, we compute a posterior distribution over that number and the boundary positions. In this section we make no authorship-attribution claim beyond the exact evidence and boundary posteriors: the tool counts distinct contiguous generating processes, and whether a process is an author is the interpretation that scholars of the book debate. The text is the book's 66 chapters in the Leningrad Codex (11,637 function-word tokens in 98 categories); 62 verse blocks serve as finer units. We apply the changepoint model of Section 3.3 with up to six processes and three settings of the prior parameter α (Appendix A tabulates every setting).

Under every prior setting and both unit sizes a single process is rejected at $\log_{10}$ Bayes factors of 41.5 to 98.2. The primary setting (chapters, $\alpha = 1/2$) gives 96.0, outside the range that shuffled chapter orders give ($p < 0.005$). Given two processes, the boundary falls after chapter 39 (posterior 0.979 to 0.994) and, with verse blocks, at Isa 40:11 (posterior 1.000 at every setting).

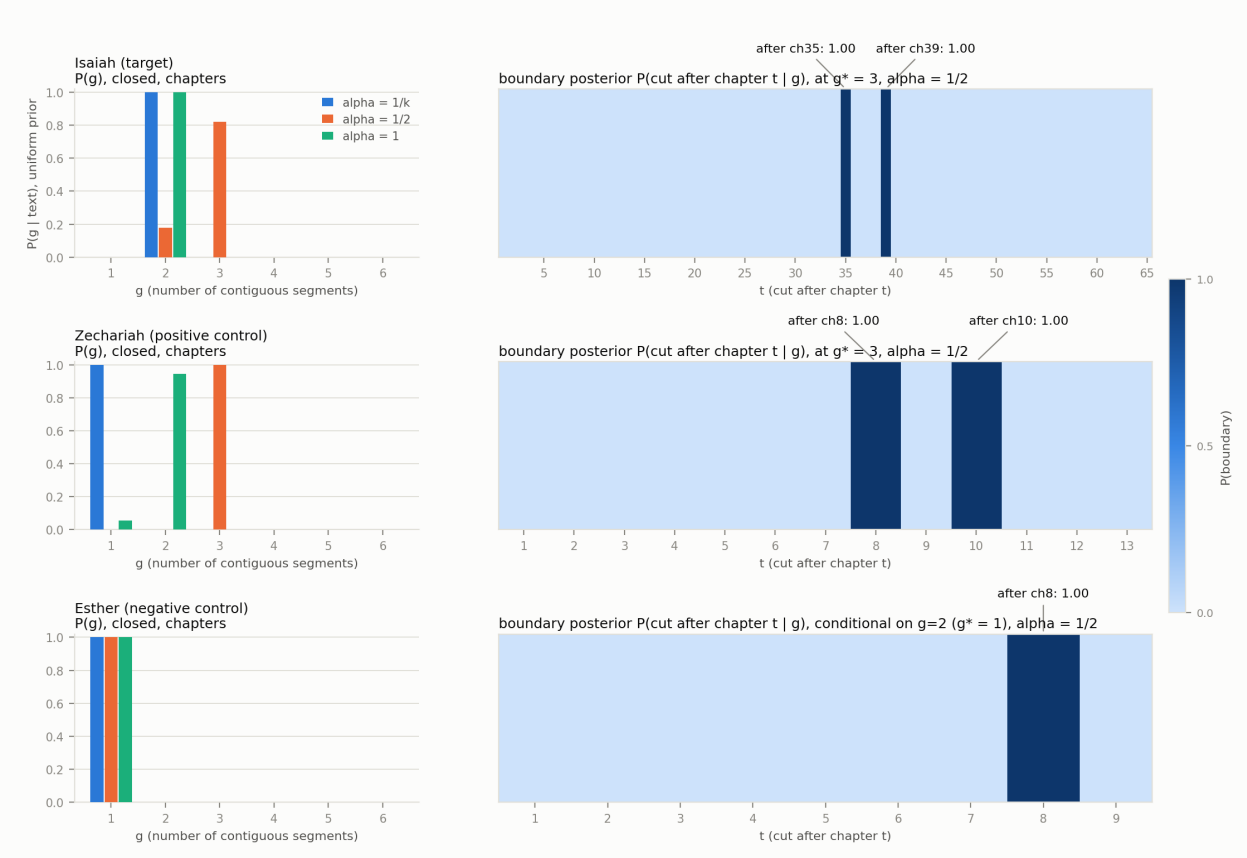

Figure 3: Number of contiguous processes and boundary positions for Isaiah and two control books (Leningrad Codex, function words, chapters). Left: $P(g)$ under the uniform prior at $\alpha = 1/k$ (blue), 1/2 (orange), and 1 (green). Right: $P(\text{cut after chapter } t \mid g)$ at $\alpha = 1/2$ for the labeled g; dark bands mark cuts. Isaiah divides at 39|40 (with three processes, also at the Hezekiah narrative after 35), with no mass at 55|56; Zechariah divides at 8|9, its count settled only at $\alpha = 1/2$; Esther is undivided (its strip shows where a forced cut falls).

Given three processes (Fig. 3), the extra boundary falls at or inside the prose Hezekiah narrative of chapters 36–39 at every setting, after chapter 35 or 38 (posterior 0.950 to 1.000; only $\alpha = 1/2$ favors three). A boundary after chapter 55 has posterior 0.000 given three and given four processes in every chapter-level setting. That zero concerns function-word rates and sharp boundaries only; it does not refute Trito-Isaiah, whose case rests substantially on content, theology, and historical setting, axes this channel set deliberately excludes, and a gradual drift within 40–66 would not present as a sharp changepoint. Chapters 40–66 analyzed alone (5,116 tokens; $\alpha = 1/2$ and 1) give the same zero, and $\log_{10}$ Bayes factors of -19.1 and -34.2 for a split at 55|56 against none (Appendix A). The power of that test is limited, however, since a synthetic change at 55|56 as large as the measured 1–39 versus 40–66 contrast (total variation about 0.19) is detected, in the median case, in only 0.62 of replicates (0.54 with verse blocks). Detection reaches eight times in ten only above a total variation of about 0.22 to 0.24, so a smaller difference within 40–66 would usually be missed; at $\alpha = 1/k$ the test has no power at this length.

For Esther, the negative control, the evidence favors a single process (Fig. 3; Table 1). Zechariah, the positive control, divides at the accepted boundary after chapter 8 with posterior 1.000 given $g \geq 2$, at a strength that depends on the prior (Table 8). The division survives when the contiguity constraint is dropped and when the narrative chapters 36–39 are excised (Appendix A).

The Great Isaiah Scroll (1QIsa^{a}) is a second witness, a copy a millennium older than the codex (Ulrich and Flint, 2010). We repeat the analysis on a tagged transcription of the scroll,[3] mapping each function-word class onto its tagset (11,444 tokens in 79 categories). The two category systems use different lemma inventories, and every agreement statement below therefore compares model outputs, the count and boundary posteriors computed separately within each witness, and never compares shared channels.

On the scroll a single process is rejected at $\log_{10}$ Bayes factors of 44.0 to 85.3 across prior settings. Given two processes, on both manuscripts the boundary lies within chapters 37–40 with posterior 1.0000 at every setting. Its most probable position is after chapter 39 on both, more weakly on the scroll (0.567 to 0.675 of the conditional mass, against 0.979 to 0.994 on the codex). With verse blocks both divide at the block edge nearest Isa 40:10/40:11. At $\alpha = 1/2$, however, the scroll's chapter-level posterior on g spreads over three to six processes, with extra boundaries in chapters 62–64, so the claim is the same boundary, never the same count. The other two settings favor two processes. Scribal drift, damage near the scroll's end, and compositional structure could each cause the spread, and we do not adjudicate among them here.

The scroll was copied in two halves of 27 columns and 33 chapters each, probably by two scribes (Popović et al., 2021), so a statistic sensitive to scribal hand would place a boundary at the seam between the halves, after chapter 33. The boundary posterior there is 0.000 given two and given three processes at every setting. The scroll's distinctive spelling (Kutscher, 1974) cannot register either, the categories being lemma-level. The seam is therefore a negative control for scribal hand and orthography; it does not show the statistic immune to transmission in general, and it says nothing about compositional structure elsewhere in the book. On the scroll too the boundary posterior after chapter 55 is 0.000 given three processes.

The computation nearest ours is the segmenter of Eisenstein and Barzilay (2008), in which each segment's multinomial is integrated against a Dirichlet prior and dynamic programming finds the most probable division into a given number of segments. Also close are penalized segmentations (Utiyama and Isahara, 2001), including Dirichlet–multinomial changepoints across an author's books (Ross, 2020), and a single change of multinomial regime among the chapters of one book (Girón et al., 2005). As far as two literature searches and a later general web search can determine, the evidence summed over all contiguous segmentations, and with it the posterior over the *number* of generating processes, had not been computed for any natural-language text.

The division after chapter 39, which matches the difference Radday (1973) found, is thus present in the function words of two manuscripts a millennium apart and unaffected by the change of scribe in the older one. Within chapters 40–66 the codex shows no boundary and the scroll none that holds across prior settings, which at the power measured above is consistent with one process there, with gradual change, or with a smaller second contrast.

[3] ETCBC Dead Sea Scrolls corpus, version 2.0 (Jacobs et al., 2026); the 84 words lacking visible consonant signs are excluded (Appendix A).

scope and treatment	$\Delta L(M_1{-}M_0)$	$\Delta L(M_{15}{-}M_1)$	$\Delta L(M_2{-}M_{15})$
Pentateuch, primary treatment, blocks	1624.0 / 1632.5 / 1625.6	943.4 / 960.6 / 965.8	−27.2 / −18.1 / −23.9
Pentateuch, one choice varied (blocks unless stated)			
split verses split	1600.5 / 1607.9 / 1598.4	954.3 / 971.3 / 975.6	−21.6 / −13.4 / −21.4
split verses dropped	1619.5 / 1626.7 / 1616.5	918.7 / 935.7 / 940.0	−22.4 / −13.7 / −20.9
revision hands to sources	1611.3 / 1622.1 / 1617.3	982.2 / 1018.4 / 1025.6	−31.2 / −20.1 / −24.6
revision hands merged	1668.5 / 1679.0 / 1673.6	866.0 / 883.8 / 886.0	−30.2 / −19.8 / −25.3
agglomerated blocks	1623.1 / 1631.6 / 1624.7	942.7 / 960.0 / 965.2	−27.9 / −18.8 / −24.5
chapters	1443.4 / 1450.8 / 1445.9	927.7 / 944.0 / 949.1	−41.9 / −33.9 / −42.8
pronoun categories removed	1482.6 / 1490.9 / 1486.6	931.8 / 948.2 / 954.3	−20.0 / −12.0 / −17.3
full vocabulary†	10086.8 / 10911.2 / 11025.3	1813.6 / 2754.2 / 3132.2	−807.7 / 32.6 / 355.5
Other scopes, primary treatment, blocks			
Hexateuch	1585.6 / 1594.3 / 1588.2	1008.8 / 1026.5 / 1033.2	−26.9 / −20.0 / −29.7
Genesis	160.0 / 178.2 / 183.5	0.0 / 0.0 / 0.0	7.9 / 11.8 / −1.9
Exodus	508.9 / 523.4 / 523.0	0.0 / 0.0 / 0.0	18.6 / 20.9 / 12.3
Leviticus	0.0 / 0.0 / 0.0	0.0 / 0.0 / 0.0	0.0 / 0.0 / 0.0
Numbers	343.2 / 347.4 / 335.5	0.0 / 0.0 / 0.0	−29.3 / −17.4 / −18.8
Deuteronomy	−16.3 / −3.3 / 3.6	0.1 / 16.4 / 20.8	−25.5 / −23.7 / −23.2
Genesis+Exodus+Numbers	1229.6 / 1239.6 / 1233.6	0.0 / 0.0 / 0.0	−24.3 / −17.0 / −24.9

Table 7: Gains in predictive score ΔL (eq. (3.7)), in nats (one $\log_{10}$ unit is 2.303 nats), between nested source maps from the Carpenter and Harford-Battersby tables: P against the rest ($M_1{-}M_0$), adding D ($M_{15}{-}M_1$), and splitting JE into J and E ($M_2{-}M_{15}$); positive values favor the finer map. Entries give values at Dirichlet parameter $\alpha = 1$, 1/2, 1/6. In every function-word Pentateuch row P and D are favored and J/E is negative (fake-partition scale: mean −27.2, maximum +33.5). Entries of 0.0 mark scopes where the two maps coincide. †Content words included.

6.3 The Pentateuchal sources

On Wellhausen's documentary hypothesis the Pentateuch is a verse-by-verse interleaving of four continuous sources (the Yahwist J, the Elohist E, the Deuteronomic D, and the Priestly P) (Wellhausen, 1885). Yoffe et al. (2023) recovered much of the Priestly/non-Priestly division of Genesis and Exodus by unsupervised clustering, and Faigenbaum-Golovin et al. (2024) separated Priestly from Deuteronomic and Deuteronomistic chapters by word frequencies. We measure, in a tournament among source maps, how well each division in the Carpenter and Harford-Battersby tables (Carpenter and Harford-Battersby, 1900) predicts held-out text (eq. (3.7)). Only the tables' divisions are compared: the tournament scores held-out predictive evidence among the named maps and discovers nothing outside the named set.

We compare four nested maps: one process (M_0); the tables' P verses against the rest (M_1); P, D, and the remainder (M_{15}); and all four sources (M_2). The unit is a block, a maximal run of consecutive verses with one label in M_2. The Pentateuch yields 526 blocks once the verses the tables mark as redactional are set aside (transcription: Appendix E).

Separating the P verses from the rest improves prediction by about 1,600 nats, 1,624 to 1,633 across the three prior settings (Table 7). The gain lies 13.2 standard deviations above the mean of ten thousand random permutations of the block labels, and 296 of the 526 blocks are individually better predicted under the split. A third process for D adds a further 943 to 966 nats, 15.9 standard deviations above its permutation mean. Both separations persist when the treatment of split and redactional verses, the block size, or the unit varies (other Pentateuch rows; Appendix A).

Nearly all D material, however, is the book of Deuteronomy. With labels permuted within each book instead, the D gain of 960.6 (at $\alpha = 1/2$) lies only 1.4 standard deviations above the relabelings (679 ± 202 nats, $p = 0.079$), and, with profiles fit within each book, adding D changes prediction by -3.1 nats. P, by contrast, keeps 992.6 nats (61%) within books and 848.5 (52%) within book and genre, beyond every permutation ($p = 0.001$). On function words D coincides with Deuteronomy and is only weakly separable from book identity even in the Hexateuch, where Joshua also carries D ($p = 0.008$ pooled, 0.07 stratified by book).

Splitting JE into J and E does not improve prediction beyond what a fake partition achieves. On the function-word categories the Pentateuch J/E difference is negative at every prior setting (-18.1 to -27.2 nats), whereas for one hundred fake four-part partitions of Esther, a single-process book (Table 1), chance alone produces gains of mean -27.2 nats and maximum $+33.5$.

Permuting the J and E labels within each book gives gains of mean -36.1 nats (standard deviation 17.8), and the observed -18.1 lies inside that null ($p = 0.17$): the tables' J/E labeling predicts no better than a random one. With synthetic differences planted between the relabeled blocks, a J/E difference of 0.10 to 0.13 in total variation (P against the rest is 0.156) clears $+33.5$ nats at least eight times in ten. The observed contrast is 0.062, and the relabeled blocks with no difference planted give 0.065. The Pentateuch J/E result is therefore a null on these channels at that measured power, and silent on the excluded axes. The documentary case for J and E rests largely on those excluded axes (doublets, content, theology, the divine names), so the null does not refute the documentary hypothesis, just as the P and D separations do not prove documents.

Part of the P and D separations may reflect genre, since P is concentrated in lists, law, and ritual, and D in oratory. Within narrative alone (419 blocks with narrative-verb share above 0.15) P still separates by $+632.3$ nats, D adds $+142.1$, and J/E ($+8.6$) stays inside the fake-partition scale. The divine names are content words, excluded by construction; also removing the sixteen pronoun-type categories closest to the classical criteria leaves the P separation at 1,482 to 1,491 nats and every Pentateuch J/E value negative (-12.0 to -20.0).

We conclude that on function words P is a distinct generating process, separated well beyond the fake-partition scale even within books and genres, and that D also separates but coincides with the book of Deuteronomy. Both findings agree with the differences that Yoffe et al. (2023) and Faigenbaum-Golovin et al. (2024) report; neither study tested J against E. J and E do not separate on these words, which were never the critics' main evidence for them, and the separations and the null are both the result. Throughout, the processes are never authors and the separations are never documents; whether P and D were documents remains the source critics' question (Schmid, 2021).

6.4 All sixty-six books

The changepoint model used for Isaiah applies unchanged to every book of the Bible. We then ask which books share a process, within each canonical section and across the whole Bible, and repeat both analyses on the oldest manuscripts.

6.4.1 Books long enough to test

A division of the reference strength, the measured contrast between Isaiah 1–39 and 40–66, is reliably recovered from roughly 4,800 function-word tokens (Section 3.4), and we call this 4,800-token power wall the threshold. Twenty-five books reach it. A stronger division can be detected in a shorter book (down to roughly 2,400 tokens) and then rests on its own Bayes factor, but a

non-detection there is uninformative. For books below 2,400 tokens, which were not scored, and for books between the two lengths with no detection, we give the token count and no result, and each such silence is stated as power, never as agreement. Besides the controls Esther and Zechariah, we analyzed five books between the two lengths. Ezra, Nehemiah, and Proverbs divide and are reported as composite; Romans and 1 Corinthians do not. Where a division falls at the boundary between Hebrew and Aramaic, as in Daniel and at Ezra 7, the change of language suffices to produce it; we note a language boundary consistent with composite origin and claim no more.

6.4.2 Results by book

In Table 8, twenty-one books are composite and nine are single at this length, meaning no seam of the reference strength is visible at that book's measured power in this channel, never proof of unity. The last block lists the remaining thirty-six books. Every one of these rows reads cannot-see, never agreement with any claim: at those lengths one process cannot be told from several.

Many of the composite results fall where source critics have placed a division. Their divisions rest largely on content, theology, historical setting, and vocabulary, which function-word counts exclude by design, so agreement and disagreement are both informative and a channel-null never refutes a content-based claim, and the comparison is model-comparison evidence, never attribution. All five books of the Torah are composite (Wellhausen, 1885; Schmid, 2021). Leviticus divides after chapter 17, where the Holiness Code begins (Carpenter and Harford-Battersby, 1900), under all three prior settings. Deuteronomy, which de Wette (1805) set apart as a source, divides after chapter 12, where its law code begins, and Job after its chapter-2 prose frame (a genre line), each at two of the three settings. Psalms has at least six processes at $\alpha = 1/2$ when up to fifteen are allowed under a uniform prior, as an anthology might, though at $\alpha = 1$ one is the most probable number. The weakest composite is 1 Kings (a $\log_{10}$ Bayes factor of 5.3, no larger than shuffled chapter orders produce).

For Ezekiel our result departs from the commentators, who treat it as the most unified of the major prophets. We find it the strongest single-language composite ($\log_{10}$ Bayes factor 363.5), with at least eleven processes at $\alpha = 1/2$ and at least seven at every prior setting when up to fifteen are allowed. Its main division falls after chapter 39, where the oracles give way to the temple vision, and may mark that change of subject rather than a second hand.

Among the books below the threshold is 2 Corinthians (2,256 function-word tokens), whose unity Semler (1776) first questioned. The view that it combines several letters is the clearest claim in New Testament scholarship that an individual book is composite. It is untested here: neither confirmed nor denied. Twenty-one of the twenty-seven New Testament books are below the threshold with no result, as is every minor prophet except Zechariah.

6.4.3 Groups of books and the whole-Bible count

Within each canonical section we also ask which books share a generating process, taking whole books as documents and each partition of them as a hypothesis. Table 9 gives the most probable partition and its number of groups K. Two controls accompany each section: whether it mixes prose and poetry, since a grouping along a genre line may count genres, and whether every book has the roughly 250 function-word tokens needed to place it, a floor measured on synthetic controls. The Catholic epistles (shortest member 102 tokens) fail the second control and enter the sums over

book	N	n	result	$\log_{10}$BF	perm. p	div. after	note	source criticism
Torah								
Genesis	–	–	composite	–	< 0.001	multiple	Table 1; ch. 11	composite: J, E, P
Exodus	12996	40	composite	167.0	< 0.005	ch. 23	all α; $g \geq 5$	composite: J, E, P
Leviticus	9897	27	composite	293.5	< 0.005	ch. 17	all α; $g \geq 9$	P; Holiness Code from 17
Numbers	11702	36	composite	124.0	< 0.005	ch. 4	all α; $g \geq 6$	composite: J, E, P
Deuteronomy	12500	34	composite	65.4	< 0.005	ch. 12	some α	D; law code from ch. 12
Former Prophets								
Joshua	7958	24	composite	110.8	< 0.005	ch. 21	all α	composite: DtrH, sources
Judges	7835	21	single	0.0	1	–	–	composite: DtrH frame
1 Samuel	10629	31	single	0.0	1	–	–	composite: DtrH, sources
2 Samuel	8502	24	single	0.0	1	–	–	composite: DtrH, sources
1 Kings	9998	22	composite	5.3	0.160	ch. 7	some α; perm.	composite: DtrH, redacted
2 Kings	9104	25	composite	46.8	< 0.005	ch. 9	some α	composite: DtrH, redacted
Latter Prophets								
Isaiah	–	–	composite	96–98	< 0.005	ch. 39	Fig. 3 (α=1/2)	chs. 1–39, 40–55, 56–66
Jeremiah	16187	52	composite	80.7	< 0.005	ch. 16	all α	composite; two editions
Ezekiel	14769	48	composite	363.5	< 0.005	ch. 39	genre line; $g \geq 11$	mostly one hand, expanded
Zechariah	–	–	composite	16.1	< 0.005	ch. 8	Fig. 3; some α	divides after ch. 8
Writings								
Psalms	13640	150	composite	98.0	< 0.005	Ps. 115	some α; $g \geq 6$	anthology, many hands
Job	6104	42	composite	18.6	< 0.005	ch. 2	genre line; some α	frame, poem; Elihu added
Proverbs	3700	31	composite	32.8	< 0.005	ch. 8	short; some α	several collections
Ezra	2656	10	composite	164.4	0.031	ch. 7	short; language	Aramaic documents; memoir
Nehemiah	4110	13	composite	64.2	0.146	ch. 3	short; all α; perm.; $g \geq 6$	memoir, Chronicler frame
Daniel	4950	12	composite	–	< 0.005	chs. 1, 7	Table 1; language	bilingual; tales|visions
1 Chronicles	7353	29	composite	183.0	< 0.005	ch. 10	genre line	Chronicler; genealogies
2 Chronicles	10365	36	composite	12.9	0.135	ch. 10	some α; perm.	Chronicler, reusing Kings
Esther	–	–	single	–	1	–	Table 1	unified novella
New Testament								
Matthew	8829	28	single	0.0	1	–	–	one evangelist using Mark
Mark	5481	16	single	0.0	1	–	–	one hand; ending added
Luke	9491	24	single	0.0	1	–	–	one hand, author of Acts
John	7707	21	single	0.0	1	–	–	one tradition, layered
Acts	8687	28	single	0.0	1	–	–	one hand, author of Luke
Revelation	4860	22	composite	6.4	0.025	ch. 3	genre line; some α	one apocalyptic hand

No result at this length ($N < 4{,}800$ with no division detected, or $N < 2{,}400$; token count only)

Romans	3391	1 Corinthians	3067	Ecclesiastes	2314	2 Corinthians	2256	Hebrews	2209
Hosea	1741	Amos	1380	Ephesians	1203	1 John	1107	Ruth	1082
Galatians	1022	Micah (Table 1)	982	Song	968	Lamentations	966	Colossians	798
Philippians	779	James	754	1 Thessalonians	747	1 Peter	729	Joel	658
Malachi	655	1 Timothy	591	Jonah	575	2 Timothy	538	Zephaniah	504
Habakkuk	465	2 Peter	455	Haggai	439	2 Thessalonians	420	Nahum	352
Titus	242	Obadiah	208	Jude	189	Philemon	172	2 John	119
3 John	102								

Table 8: Every book of the Bible under the changepoint model. Columns: function-word tokens N; units n; result (single: no division at this length); $\log_{10}$ Bayes factor of the best-supported number of processes against one; its chapter-permutation p at $\alpha = 1/2$ (<: exceeds all 200 or 1000 shuffles; 1 if single); chapter before the main division; note; position in source criticism (DtrH: Deuteronomistic History). Notes: all or some α, composite at every prior setting or only some; genre line, the division coincides with a change of genre; $g \geq m$, at least m processes at $\alpha = 1/2$ (95% credible, fifteen allowed); perm., no larger than shuffled chapter orders give; short, below 4,800 tokens; language, the Hebrew|Aramaic boundary; Table 1 or Fig. 3, result from there.

section	D	$K_{1/2}$	K_1	partition at $\alpha = 1/2$	controls (genre strata; smallest book)
Torah	5	4	4	{Gen} {Exod, Num} {Lev} {Deut}	$P(K{=}4) = 1.0$; one genre; all books above the floor
Former Prophets	6	3	2	{Josh} {Judg, 1Sam, 2Sam} {1Kgs, 2Kgs}	$P(K{=}3) = 1.0$; prior-dependent; one genre
Major Prophets	3	3	3	{Isa} {Jer} {Ezek}	$P(K{=}3) = 1.0$; one genre
The Twelve	12	3	2	{Hos} {Joel, Amos, Jonah, Zeph, Hag, Zech, Mal} {Obad, Mic, Nah, Hab}	$P(K{=}3) = 0.99945$; mixed genre, and the partition follows the poetry/prose line; thin books (Obad 208 < 250)
Writings	13	8	5	{Ps} {Prov} {Job} {Song, Lam} {Ruth} {Eccl} {Esth, Neh, 1Chr, 2Chr} {Dan, Ezra}	$P(K{=}8) = 0.751$, 90% interval $[7, 8]$; mixed genre (3 strata), and the partition never crosses the poetry/prose line; prior-dependent
Gospels + Acts	5	3	3	{Mt, Mk, Lk} {Jn} {Acts}	$P(K{=}3) = 1.0$; one genre; smallest book 5481
Pauline corpus	13	2	1	{Rom, 1Cor, 2Cor, Gal} {the other nine}	$P(K{=}2) = 1.0$; prior-dependent; thin books (Phlm 172, Tit 242 < 250)
Catholic epistles	7	indeterminate		(not partitioned)	thinnest book 102 < 250; $K > 1$ cannot be resolved at this length

Table 9: Which books of each canonical section share a generating process. D is the number of books; $K_{1/2}$ and K_1 are the most probable numbers of processes at $\alpha = 1/2$ and $\alpha = 1$ (prior-dependent: the two differ). Then the most probable partition at $\alpha = 1/2$ with its posterior probability, the section's genre make-up (a partition along a genre line may count genres), and its smallest book against the 250-token floor below which a book cannot be placed. These counts are the addends of the section sum discussed in the text.

sections below as an indeterminate placeholder of two processes; Hebrews and Revelation enter as one process each.

The Torah, a section of uniform genre, resolves into four processes at both prior settings, but interleaved Pentateuchal sources (Section 6.3) cannot be separated by a partition of whole books. Judges and the books of Samuel, each single at this length, form one process (without Joshua); Kings divides internally and stands apart at $\alpha = 1/2$, rejoining them at $\alpha = 1$. In source criticism Joshua through Kings form one Deuteronomistic History, whose editorial layer recent word-frequency work treats as one corpus (Faigenbaum-Golovin et al., 2024). One editorial hand is consistent with the uniform rates of Judges and Samuel, though the divisions inside Kings suggest it stops short of the whole history.

The synoptic gospels share one process and John stands apart, as the dependence of Matthew and Luke on Mark predicts (Holtzmann, 1863; Streeter, 1924), since copied text keeps its function-word rates. Acts, however, separates from Luke ($P(K{=}3) = 1.0$), and since both are narrative, the genre control, which distinguishes prose from poetry, does not account for the split. Common authorship of Luke and Acts is the view of most commentators and of a recent stylometric study (Savoy, 2026). We conclude no more than that their function-word processes differ, even if one hand wrote both. The Pauline letters form two processes at $\alpha = 1/2$, the core being exactly the four Hauptbriefe of Baur (1845), and one process at $\alpha = 1$; Philemon (172 tokens) is too short to place securely.

Summed over sections, the counts of Table 9 come to 16 to 21 in the Old Testament and 8 to 9 in the New at the two prior settings, 24 to 30 in all. That sum assumes no profile is shared

across sections, which the comparisons below do not support. Taking all sixty-six books at once, we merge books or groups of books, the most favored pair first, as long as the Bayes factor from eq. (3.2) favors a shared profile over separate ones. We then re-partition each merged group of up to thirteen books by full enumeration (Appendix C). Several of the resulting profiles span sections at $\alpha = 1/2$: Genesis groups with Judges, Samuel, Jonah, Ruth, and Haggai; Isaiah with Micah, Nahum, Habakkuk, and Obadiah; the synoptic gospels with James; the Pastoral letters and 2 Thessalonians with the shorter Catholic letters; and 1 Peter with Pauline letters. Merging whole books across the canon thus leaves 30 to 32 distinguishable function-word profiles at $\alpha = 1/2$ and 23 to 24 at $\alpha = 1$ when each comparison uses only the word categories present in the books compared. With the full Hebrew or Greek category list in every comparison, the counts are 19 and 13. These are counts of profiles under stated priors and vocabularies, obtained greedily in the merge order and exhaustively within each merged group. Each is a count at this channel's resolution and never an author count or a bound on the number of sources, and the divisions within books (Table 8) are not added to it. Over the wider band from $\alpha = 1/k$ to 5 the section sums run from 32 down to 10, so the count is band-conditional as well as vocabulary-conditional.

6.4.4 The older manuscripts

Because the main analysis uses one medieval codex and one modern edition, we repeated it unchanged on the oldest available witnesses: the Samaritan Pentateuch, the two Isaiah scrolls from Qumran, the Septuagint, Codex Sinaiticus, the Pauline papyrus P46 (about 200 CE), a Masoretic edition of the Aleppo line, and for Jeremiah a Hebrew text aligned to the shorter Greek edition.[4] Each witness is analyzed on its own word categories; we therefore compare only its results with the main analysis.

Of the 30 books with a result, 26 survive in a witness earlier than 500 CE, and 25 of the 26 checkable repeat their classification on at least one line of transmission. That 25 of 26 is the most favorable of these tallies. On each book's earliest witness alone the count is 19 agreeing, 1 partial (class kept, division moved), 5 diverging and 1 untestable. Six books replicate on an edition line only (the Aleppo-line edition of about 930 CE in every case), a test of the edition more than of age. Isaiah's division after chapter 39 (Duhm, 1892) recurs on the Great Isaiah Scroll, copied roughly 1,133 years before the codex (Section 6.2). The Torah, gospel, Acts, and Pauline results likewise recur on the Samaritan Pentateuch, Sinaiticus, and P46, with small differences (Appendix C).

The one replication failure is 1 Kings, the weakest composite, single on both lines that carry it. Four other divergences occur only in Greek translation; Daniel's is expected, since translation removes a division that was a change of language (Appendix C). Where an early same-language witness survives, the classification thus appears to belong to the transmitted text, centuries older than the edition we analyzed.

7 Discussion and conclusions

Treating a disputed text as a mixture makes each division a hypothesis with its own evidence, which we computed in closed form for every Federalist paper, Shakespeare play, and sufficiently long biblical book. The joint Federalist papers receive a Hamilton share comparable to, or below,

[4] Isaiah scrolls: Jacobs et al. (2026), after Ulrich and Flint (2010); Septuagint: Codex Vaticanus; Jeremiah: the Computer Assisted Tools for Septuagint Studies (CATSS) aligned text (University of Pennsylvania).

that of a typical paper by Madison alone. Six plays show a second dramatist, and four long enough to test show none. Isaiah divides after chapter 39 in two manuscripts a millennium apart. In the Pentateuch, P separates from the rest even within books and genres, while D's separation coincides with the book of Deuteronomy. Across the whole canon the books merge into roughly a dozen to thirty function-word profiles, depending on prior and vocabulary, and several profiles span canonical sections; the count is of profiles, never of authors.

Most of the results confirm the scholarly consensus. In *Henry VIII*, however, the hand Spedding (1850) and Hoy (1956) give to Fletcher is closer to Middleton's under both count models, though two Jacobean dramatists may look alike on function words. Ezekiel, which commentators treat as the most unified major prophet, is the most sharply divided single-language book, its main division falling at the change from oracles to temple vision. Acts stands apart from Luke, though most commentators and a stylometric study (Savoy, 2026) assign both to one author. The word counts give no support, at the power these texts allow, to dividing Isaiah after chapter 55 (Duhm, 1892) or to separating J from E, two cases built on content and theology, which function words do not carry.

The main analysis treats function words as independent draws, an assumption Mosteller and Wallace (1964) rejected because such counts overdisperse; Bayes-factor magnitudes are relative to that assumption. Under the hierarchical model the joint papers stay Madison-dominant and four of the six Shakespeare detections remain. Under it, however, No. 55's preference for a blend disappears, no single act of *2* or *3 Henry VI* stays decisive, three of the four "none" results become unresolved, and one of fifteen solo plays falls on Marlowe's side. In both models, ninety-percent intervals for the blend weight miss a known true value well over one time in ten, so the weights order texts without measuring fractions of authorship. A collaborator without undisputed work has no profile and cannot be detected. In only one play would a tenth of the verse by any of the comparison dramatists change the classification, and help written in prose is outside the test. Over half the biblical books and most scenes are too short to test, and a detected division may mark a change of genre or language.

A count model with calibrated weight intervals would turn the weights into measured fractions of authorship. A second securely attributed solo play for Kyd, Nashe, or Wilkins would make their profiles usable, though none is known to survive. More generally, the evidence integrals apply wherever a division has been argued on literary or historical grounds and the text is long enough to test.

Data and text sources

All texts analyzed are publicly available. The principal ones are the Project Gutenberg text of the *Federalist*, EEBO-TCP transcriptions of early modern plays, the Open Scriptures morphology of the Westminster Leningrad Codex, the ETCBC transcription of the Dead Sea Scrolls (Jacobs et al., 2026), the tagged SBL Greek New Testament, and the source tables of Carpenter and Harford-Battersby (1900); the older biblical manuscripts are named where they are used (Section 6.4). Appendix E describes the word lists, the transcription of the source tables and the computations.

Acknowledgments

Funding for this project was provided by Anthropic. This is not an official Anthropic project and results and views are not endorsed by Anthropic.

A Changepoint configurations

Table 10 gives the posterior on the number of processes in Isaiah (Leningrad Codex) for the six configurations of Section 6.2 and six more that count all lemmas; a geometric prior on g changes no row's most probable count.

Verse blocks are built separately on the two Isaiah manuscripts and their edges differ: near chapter 40 the scroll has no block edge at Isa 40:11 and the codex none at 40:10. Given two processes the boundary falls after 40:10 on the scroll (posterior 0.979 to 0.988 across prior settings) and after 40:11 on the codex (1.000 at each); the placements agree to within the blocks' resolution. With verse blocks at $\alpha = 1/2$ the scroll's count posterior concentrates on five processes (0.979); with chapters it spreads over three to six. On the codex, scoring all 2^{20} two-label assignments of the text regrouped into 20 blocks, without the contiguity constraint, leaves 99.5 to 100% of the posterior mass on contiguous divisions. With the narrative chapters 36–39 excised, the remaining 62 poetic chapters still divide at the excision ($\log_{10}$ Bayes factors 38.2 to 91.2 against one process). The control for the direct test of chapters 40–66 (Section 6.2), chapters 27–53 thinned to the same length, divides after chapter 39 in 20 of 20 replicates at $\alpha = 1/2$ and at 1, in none at $1/k$.

From the scroll's counts we exclude 84 words with no visible consonant signs (76 fully reconstructed and 8 sign-less, of 22,776 tagged), since damaged passages are restored from the Masoretic text. With these words included the scroll's conclusions hold: two processes at two of three prior settings, the boundary within chapters 37–40 at posterior 1.0000, and zero boundary posterior after chapters 33 and 55; the posteriors shift little, and not uniformly toward the codex.

In Daniel the language switch at 7:28 is a verse-block edge, recovered at posterior 1.0000 in every block-level function-word configuration. The switch at 2:4 lies inside a block, and the inferred cut moves to a neighboring edge: the preceding one (1:17, posterior at least 0.9991) on the function-word list and the following one (2:13) on the full-lemma list at two of three settings.

In Genesis the two-process cut falls after chapter 11, 41 or 1 at $\alpha = 1/2$, 1 or $1/k$ respectively. With verse blocks the corresponding cuts fall at 11:29 and 41:57, and the $\alpha = 1/k$ cut spreads over 1:21 and 2:7. Given three processes the cut after chapter 11 is stable and the second moves with α: after chapter 9 at $1/k$, after chapter 41 at 1/2 and 1, with the rest of its posterior mass after chapter 33. On the topic-sensitive full-lemma list one process is rejected for Daniel at $\log_{10}$ Bayes factors of 818.1 to 1755.2 and for Genesis at 38.6 to 666.4; Genesis's most probable count varies between two and four with α. A geometric prior on g leaves the most probable count unchanged in all 48 configurations of the four books (Section 3.5).

Micah stays at one process in every function-word configuration, so the following cut positions, computed given two or three processes, are not evidence of division. Given two processes the chapter-level cut falls after chapter 4 at every α, a position in neither commentary division. Given three, the cuts fall after chapter 6 and, depending on α, after chapter 4 or 5; the break after chapter 5, common to both divisions, has posterior 1.0000, 0.1034 and 0.3225 at $\alpha = 1/k$, 1/2 and 1. On the full-lemma list two processes are preferred at one prior setting for both units ($\log_{10}$ Bayes factors 12.2 and 12.5).

unit	word list	α	n	k	$P(g{=}1)$	$P(g{=}2)$	$P(g{=}3)$	$\log_{10}\mathrm{BF}(g^*{:}1)$
chapter	closed	$1/k$	66	98	0.000	1.000	0.000	41.5
chapter	closed	1/2	66	98	0.000	0.178	0.822	96.0
chapter	closed	1	66	98	0.000	1.000	0.000	69.1
block	closed	$1/k$	62	98	0.000	1.000	0.000	44.3
block	closed	1/2	62	98	0.000	1.000	0.000	98.2
block	closed	1	62	98	0.000	1.000	0.000	72.0
chapter	full†	$1/k$	66	2570	1.000	0.000	0.000	0.0
chapter	full†	1/2	66	2570	0.000	1.000	0.000	32.2
chapter	full†	1	66	2570	1.000	0.000	0.000	0.0
block	full†	$1/k$	62	2570	1.000	0.000	0.000	0.0
block	full†	1/2	62	2570	0.000	1.000	0.000	20.9
block	full†	1	62	2570	1.000	0.000	0.000	0.0

Table 10: Posterior on the number g of contiguous processes in Isaiah (Leningrad Codex) by configuration: unit (66 chapters, or 62 verse blocks of at least 400 tokens of running text), word list (closed: the function-word categories; full: every lemma, content words included) and Dirichlet prior parameter α. Columns give units n, categories k, the posterior on $g = 1, 2, 3$ under a uniform prior on $g \leq 6$ ($P(g \geq 4)$ is 0.000 throughout), and the $\log_{10}$ Bayes factor of the most probable count g^* against a single process. One process is rejected in every function-word row; in the one three-process row the additional boundary is after chapter 35. †Full-lemma counts also reflect topic and genre, and the conclusions therefore rest on the function-word rows.

In the primary rows of Table 7 the verses that the Carpenter and Harford-Battersby tables mark as redactional are set aside and a verse divided between sources carries one label. The revision-hand rows treat the redactional verses as a separate source or give them the neighboring source's label; the split-verse rows give a divided verse its last part's label or drop it. The agglomerated row merges consecutive same-label blocks until each has at least 200 tokens, the chapter row uses chapters as units, and the pronoun row omits the sixteen pronoun-type categories. On synthetic texts generated from known two- and four-source maps the true map is favored over one process by +565.5 and +819.8 nats; the four-source text reuses the tables' J and E blocks, and its J/E split scores −12.2 nats.

The within-narrative analysis keeps the 419 blocks whose share of narrative verb forms, a mechanical proxy for genre, exceeds 0.15. At threshold 0.1 (438 blocks) the P, D and J/E margins are +1006.0, +330.4 and +12.6 nats, and at 0.25 (359 blocks) +571.6, +106.4 and +14.3. The best-scoring free two-way labeling of Genesis regrouped into twenty blocks lies at Hamming distance 2 from the tables' P map (purity 0.65), and the best free four-way labeling of twelve blocks lies at Hamming distance 5 from the tables' four-source map (purity 0.58). Run per book on chapters with up to four processes allowed, the changepoint model prefers three for Genesis and reaches the four allowed for Exodus, Leviticus, Numbers, Deuteronomy, and Joshua (Table 8 allows fifteen); such counts do not bear on the tables' interleaved maps.

B Shakespeare decision rule and reference sets

This appendix gives the classification rule, the census under the hierarchical model, two profile-free analyses, and the *Henry VI* reference sets of Sections 5.2 and 5.3.

B.1 The classification rule

For a play and a pair of profiles, M and C_{reg} are defined in Section 5.2; the solo plays that set C_{reg} are withheld from the Shakespeare profile when scored, and the margin is the standard deviation of M over them. A second hand is reported only if M exceeds C_{reg} by the margin and at least one act or sufficiently long scene is decisive. An act is decisive when its whole band of $\log_{10}$ Bayes factors, over the full profiles and every profile with one play removed, favors the comparison dramatist by more than 2 and by more than any act of another solo play of the period. A scene is long enough if reference scenes of the two dramatists separate at its length. The play under test never sets its own ceiling, margin, or act threshold; applied to the fifteen solo plays, each withheld in turn, the rule reports a second hand in none.

The absence of a second hand is reported only where a planted one changes the classification (Section 3.4). Verse from a withheld play of each comparison dramatist of the period, 0.20 of the play's verse, is planted as one stretch, dispersed over every scene, or as one whole act. For every dramatist and mode it must move the play out of the "none" class under this rule in at least 0.8 of 96 replicates. We measured this for thirteen plays. Eleven are the plays without a gain above any ceiling that pass a cruder form of the test, in which a fifth simulated from the comparison profile itself need only lift M above C_{reg} in 0.8 of 24 trials. The other two, *Richard II* and *Coriolanus*, are hosts of the length and period of *3 Henry VI* and *Henry VIII*. Five pass, the four classed "none" and *Richard II*, unresolved because its gain is not localized (Table 5); in the other eight the play too often stays "none" with a planted fifth, of Greene's or Marlowe's verse especially, frequently with no blend gain at all. With a whole act of Peele's planted, the play is classed "had help" in 0.64 to 1.00 of replicates in eight of the eleven Elizabethan hosts, and with a Fletcher act in *The Tempest* in 0.77. For Greene's or Marlowe's verse the rate is 0.00 to 0.64, and for a dispersed fifth almost zero, except Fletcher's in *The Tempest* (0.39). Plays that fail it without showing a second hand are unresolved, as are plays with a gain above a ceiling but no decisive act for any one pair.

B.2 The census under the hierarchical model

Under the hierarchical model of Section 3.2 the concentrations are estimated from each side's solo plays, each left out in turn, and every threshold and the power are remeasured in the same way. The ceilings become 0.85 (Peele), 2.83 (Marlowe), 1.28 (Greene), −0.70 (Fletcher), and 4.64 (Middleton), and the census reads four had help, one none, and thirty-six unresolved (Table 11 for the resolved plays of Table 5). Of the fifteen solo plays, withheld in turn, none is classed "had help," but *Henry V* falls on Marlowe's side (Table 11), as no solo play does under word independence. Throughout, w is the weight of the Shakespeare profile in the best-fitting blend. On 576 constructed mixtures its 90 percent interval covers the true list-token share in 41 percent of cases under word independence and 49 percent under the hierarchical model, so w orders the plays but is not a calibrated fraction of the text.

B.3 Two profile-free analyses

In the first analysis we measure how much function-word rates vary between a play's acts, as a z-score against 300 homogeneous resamplings. Every play with several acts is heterogeneous by this measure (z from 6.4 to 20.5; $p = 0.0$ on 40 of the 41), yet on four solo plays a synthetic second-author share of 0.1, 0.2, or 0.3 is detected at rate 0.0, since solo plays themselves reach $z \approx 20.4$.

play	word-independence model profile	M	C_{reg}	power	result	hierarchical model profile	M	C_{reg}	power	result
3 Henry VI	Marlowe	53.47	7.58	–	had help: act 1	Marlowe	16.16	2.83	–	gain not localized
2 Henry VI	Marlowe	27.88	7.58	–	had help: act 1	Marlowe	8.77	2.83	–	gain not localized
Edward III	Greene	46.59	1.22	–	had help: acts 3, 5	Marlowe	8.19	2.83	–	had help: act 3
Titus Andronicus	Peele	27.98	1.92	–	had help: act 3	Marlowe	4.96	2.83	–	had help: act 1
Henry VIII	Middleton	62.55	19.58	–	had help: acts 4, 5, prologue	Middleton	9.82	4.64	–	had help: act 4, prologue
Two Noble Kinsmen	Middleton	56.38	19.58	–	had help: acts 2, 4, epilogue	Middleton	10.63	4.64	–	had help: acts 2, 4, epilogue
King John	Peele	−0.10	1.92	0.84	none	Peele	0.26	0.85	0.98	none
Midsummer Night's Dream	Greene	1.17	1.22	0.93	none	Greene	1.28	1.12	0.79	power too low
Richard III	Marlowe	1.68	7.58	0.86	none	Peele	0.85	0.57	1.00	gain not localized
Tempest	Fletcher	0.64	2.36	0.90	none	Middleton	−0.79	4.64	0.31	power too low
Henry V (solo)	Marlowe	6.09	7.58	–	gain not localized	Marlowe	−0.04	2.83	–	leans Marlowe

Table 11: The six detections, the four plays classed "none," and the solo *Henry V*, scored by the same rule under the word-independence model (as in Table 5) and under the hierarchical model, with ceilings, margins, and power remeasured under it. Profile: the comparison dramatist that decides the classification under each model; M, C_{reg}, and power as in Table 5, with C_{reg} set without the play itself where it is one of the solo plays. *Henry V*: Shakespeare-against-Marlowe $\log_{10}$ Bayes factor -8.91 at the full profiles (its band crosses zero), $w = 0.23$ $[0.16, 0.30]$.

In the second we partition whole plays by function-word counts alone within each of seven sets (the Folio comedies, histories, tragedies and Roman plays, the two history tetralogies, and all 36 Folio plays). We compute the posterior on the number of groups K by enumeration for ten plays or fewer and by Gibbs sampling otherwise, under three partition priors. In every set the whole 90 percent interval lies at K=1. Two controls suggest that this reflects low resolution at whole-play scale: a solo play split in two is rejoined at rate 1.0 on all six trials, but texts joined from two dramatists' plays are separated at rate only 0.125, one pair in eight (*Hamlet* with Fletcher-profile text). Function-word rates of whole plays may differ more with genre and period than with author.

B.4 Henry VI reference sets

When the nine reference plays are withheld in turn, the 142-word list classifies eight correctly, and three alternative lists seven, seven, and eight. Table 12 summarizes the three earlier reference-set designs.

In the third design the withheld *Edward II* falls on the correct side only when *Dido* and *The Jew of Malta*, whose Marlowe authorship is itself contested, are both in the Marlowe profile: its Shakespeare-against-Marlowe Bayes factor is -7.0 with the full six-play profile, but $+23.2$ without *Dido* and $+78.8$ without *The Jew of Malta*. *The Jew of Malta* in turn falls Marlowe-side only with *Edward II* present. The reference plays' Bayes factors are more sensitive to genre than to author: in the first design, removing the histories from the Shakespeare profile moved *Richard II* and *Richard III* about 60 $\log_{10}$ units toward Marlowe. In the third, Bayes factors from pooled and from genre-matched profiles differed by $[-15.8, +76.8]$, against an authorship margin of about 7

stage	design	reference plays misclassified	outcome
1	eleven reference plays, three word lists, each play withheld in turn	*Richard II* wrong-sign at -49.0; *Edward II* and *Faustus*-A wrong-side under the full vocabulary; *Richard III* and *A Midsummer Night's Dream* fail the requirement that the posterior of w concentrate near the pure-author value	fail
2	verse lines only, stratified by register; word list chosen by genre-stratified comparison; Marlowe bands from leaving one play out	Shakespeare plays all correct, 8 of 8; *Edward II* $+158.5$, band $[+137.2, +200.6]$, and *Faustus*-A $+26.0$, band $[+13.1, +61.5]$, wholly wrong-side; *Richard II* fragile when scenes are left out in turn, minimum margin -8.9	fail
3	*Dido* and *The Jew of Malta* added as a labeled second group; genre-matched comparison profiles; register-robust word lists rebuilt for each withheld play	*Edward II* matched band $[-7.0, +78.8]$, not wholly Marlowe-side	discarded

Table 12: Three earlier reference-set designs for the *Henry VI* comparison of Section 5.3, with the withheld reference plays that each misclassified or classified only marginally. Entries are $\log_{10}$ Bayes factors for the Shakespeare profile against the Marlowe profile: a Shakespeare play is classified correctly when positive, a Marlowe play when negative; bracketed bands are ranges over the full Marlowe profile and its variants with one play removed. The last column gives the outcome: the first two designs were revised after these failures and the third was discarded.

on *Edward II*. With about 232,500 verse tokens over fourteen plays and about 4,900 function-word tokens in the *Edward II* comparison, a genre shift of tens of $\log_{10}$ units and an authorship signal in single digits could not be separated.

When each act of *The Spanish Tragedy* (the whole Kyd profile) is withheld and compared with Shakespeare, acts 1, 2, and 4 give $+12.1$, $+7.9$, and $+4.8$ for Kyd ($N = 942$, 681, 903) and act 3 gives -5.6, against Kyd ($N = 1956$). With one of its own acts classified against Kyd, the profile serves only as a check.

Reference scenes of Shakespeare and Marlowe separate cleanly only from $N = 400$ function-word tokens, and 73 of the 77 *Henry VI* scenes are shorter. At these lengths a solo Shakespeare play would give a decisive result on about 39% of scenes and a solo Marlowe play on about 19%, with about 8% and 1% decisive the wrong way. Among the 178 reference scenes themselves, eight are decisive for the wrong profile, a rate of 0.045. Of the *Henry VI* scenes 23 are decisive (Fig. 2): seven Shakespeare-side and none Marlowe-side in Part One, six and none in Part Two, three and seven in Part Three.

C Whole-Bible checks and manuscript replication

This appendix reports the checks and manuscript-by-manuscript replication behind the survey of all sixty-six books (Section 6.4).

C.1 Short books and Job's second division

We checked three books under 2,400 function-word tokens, which Table 8 lists unscored; none enters the survey's counts. For Hosea the evidence favors two processes at every prior setting ($\log_{10}$ Bayes factors 23.0, 11.0, 13.7 at $\alpha = 1/2$, 1, $1/k$). The division, after chapter 3 where prose gives way to poetry, may mark that change of genre. Lamentations is composite, with three processes at $\alpha = 1/2$ ($\log_{10}$ Bayes factor 16.7), two at $\alpha = 1$ (7.4), and no division at $\alpha = 1/k$; in Ecclesiastes two processes are favored only at $\alpha = 1/2$, and marginally (7.2).

In Job the division after chapter 2 sets off the prose frame. A second division, after chapter 37 where the Elihu speeches end, appears only in the three-process solution ($\log_{10}$ Bayes factor 18.6 in Table 8, uniform prior). Under the geometric prior on the number of processes, that solution has posterior probability 0.47 and the most probable number is two, at a $\log_{10}$ Bayes factor of 18.3.

C.2 Coarser groupings

In a first grouping coarser than Table 9 we partition all canonical sections of both languages together using rates of six part-of-speech classes: $K = 7$ blocks ($P(K{=}7) = 0.99999$), four in the Old Testament and three in the New, none spanning both. In the second we group each Testament's five sections using function-word counts, and every reported partition has posterior probability 1.0. The Old Testament forms {Torah, Former Prophets} {Major Prophets, the Twelve} {Writings} at $\alpha = 1$ and four blocks at $\alpha = 1/2$, where Torah and Former Prophets separate. The New Testament forms {Gospels and Acts} {Pauline, Catholic, Hebrews} {Apocalypse} at either setting; both coarser groupings follow genre and language.

C.3 The whole-Bible count

Under a uniform prior over each canonical section's partitions the sum over sections (Section 6.4) is sharply determined: its posterior is $P(30) = 0.751$, $P(29) = 0.169$, and $P(31) = 0.080$ at $\alpha = 1/2$ and $P(24) = 0.999$ at $\alpha = 1$, below the prior's mode of 34 (90% interval $[30, 38]$). The merged count starts from the sixty-six books' count vectors (39 on Hebrew categories, 27 on Greek, never compared across Testaments). Two groups with pooled counts a and b are compared by $\log_{10}\{Z_\alpha(a+b)/[Z_\alpha(a)\,Z_\alpha(b)]\}$ from eq. (3.2) in closed form; the highest-scoring pair is joined, and the step repeated while the highest score is positive. Requiring a Bayes factor of ten instead changes the counts by at most one. Each merged group of at most thirteen books is then replaced by its most probable partition under the posterior over all its set partitions, enumerated in full (uniform prior), which splits two groups at $\alpha = 1/2$. The same check on the union of neighboring final groups moves one boundary inside the Pauline corpus without changing the count. The New Testament groups of fourteen to twenty books that form at $\alpha = 1$, or when every comparison uses the full Greek list, exceed thirteen and stay as merged; only in the analysis that gives 32 is every group re-partitioned.

C.4 Replication by book and manuscript

We compare each book that has a result with every older text line preserving enough of it. A translation shares no function words with its source. We therefore designed 57 comparison cells with every cross-language cell fenced to seam locations and cluster topology only, as model-output comparisons and never shared-channel statistics, and never attribution. Of the 57, 44 agree with

the survey's classification, 5 keep the class with a division moved, 6 diverge, and 2 cannot be made (the Greek Psalms and Jeremiah differ in chapter numbering and order). However, three of the 44 (Romans and 1 Corinthians on Sinaiticus, 1 Corinthians on P46) are too short to test on both sides and omitted from the replication counts.

Six comparisons diverge: 1 Kings on both the Aleppo-line edition and the Greek, and Judges, 2 Chronicles, Job, and Daniel on the Greek only, where for Judges the evidence favors two processes at a $\log_{10}$ Bayes factor of 2.3. On the Aleppo-line edition we analyze the verses whose consonantal text matches the Leningrad Codex (62.8 to 94.4% of verses per book; 91.7% for 1 Kings). Of the 317 book copies on Dead Sea scrolls (Jacobs et al., 2026), 316 (all but the Great Isaiah Scroll) fall below every threshold (median 41 function-word tokens). The second Isaiah scroll, 1Q8, comes nearest (2,345 tokens against 2,400) but is not among the 57 comparisons.

All five Torah books are composite on the Samaritan text with divisions identical or one chapter away, and the same four-process grouping holds at $\alpha = 1$. Sinaiticus reproduces the gospel, Acts, and two-process Pauline results. Counted once each by its best line, the 26 books with a result and an early manuscript replicate on an ancient copy in the original language (12), on an ancient translation (7), on an edition line only (6), or nowhere (1, 1 Kings). The Pauline letters' split at $\alpha = 1/2$ recurs on Sinaiticus with 2 Corinthians outside the Hauptbriefe core. On P46 it recurs with Philippians (653 tokens there) inside the core and 1 Thessalonians below the 250-token floor. Adding Hebrews (outside the survey's Pauline set) to the P46 letters isolates it at $P(K{=}2) = 0.99999$; that analysis enters no count.

C.5 Jeremiah's two editions

The Greek Jeremiah is shorter and differently ordered. Through the CATSS alignment we projected the survey's Hebrew text onto its extent and order. Jeremiah's three divisions survive removal of the Masoretic expansion: none lies inside it, and the principal one stays at chapter 16. Only the edges of the oracles against the nations move, from 45|46 in the Masoretic order to 25|26 and 31|32 in the Greek, and they also separate poetry from prose. Removing the expansion (14.1% of the words) raises the posterior probability of the chapter-16 boundary from 0.878 to 0.998, as if the expansion had blurred an older division. Because this projection reuses the survey's own text, it tests the edition, and Jeremiah stays among the six edition-line replications.

D Federalist robustness and calibration

Against profiles estimated from all 65 papers of known authorship, the pure-author comparison under the inherited 70-word list has the correct sign for 51 of 51 Hamilton and 14 of 14 Madison papers. That fit is in-sample, and we do not rest the accuracy statement on it; the statement is the leave-one-paper-out count (Table 2).

No. 20 (posterior mean 0.1183 under the inherited list, Table 3) remains Madison-dominant in five variant analyses. Dirichlet smoothing of the profiles at $\alpha_0 = 1/10$ and 10 in place of 1 gives means of 0.1246 and 0.0968, and priors on w from Beta$(1/2, 1/2)$ to Beta$(3, 2)$ give 0.0772 to 0.2211. Re-estimating the profiles 24 times, leaving out each Madison paper in turn or a block of five or six Hamilton papers, keeps the mean within $[0.0849, 0.1837]$. Replacing the inherited 70 words by the training papers' 70 most frequent gives 0.1129, and adding a seventy-first category for all other tokens ($k = 71$) gives 0.0964. For No. 55 under the inherited list, the same 24 refits

spread the pure-author comparison over $[-0.537, +1.914]$ while the blend's margin over either pure author never falls below $+1.539$.

Under the inherited list the two halves of No. 20 have Hamilton shares of 0.3552 and 0.0887; split into halves and thirds of a few hundred function-word tokens each, Nos. 18 and 19 vary internally with less spread. The Hamilton-weight means for No. 18 are 0.0670 and 0.0584 by halves and 0.0630 to 0.1107 by thirds, against 0.0368 for the whole paper. For No. 19 they are 0.0755 and 0.1318, and 0.0720 to 0.2239, against 0.0579.

To measure the power of the blend-versus-pure comparison we joined opening segments of a Hamilton and a Madison paper, both excluded from the profiles, at a fixed token ratio in each of ten pairs, under the inherited and closed-class lists. At an even split (N = 984–4094) the blend is preferred to the better pure model by more than 10^2 in twenty of twenty trials. The posterior mean of w then recovers the true share with root-mean-square error 0.11 (largest miss 0.24). At three-to-one splits in either direction the factor is reached in 24 of 40 trials, and the posterior means are then biased toward equal shares by 0.06–0.15. Repeated under all three lists, the even-split trials reach the factor in 3, 9, and 7 of ten under the hierarchical model (inherited, closed-class, top-300), against 10, 10, and 9 under word independence.

Jay's five papers (Nos. 2–5 and 64) were not used in any profile or word list, and neither pure model represents a third author. Under the inherited list four of them have $\log_{10}$ Bayes factors below -4 (Madison's side), and for No. 5 the blend is preferred to the better pure model ($\log_{10}$ Bayes factor $+1.840$). Over all 85 papers at equal prior odds, the pure models give an expected 52.0, 52.1, and 55.0 Hamilton papers under the three lists, and 51.5, 51.0, and 51.0 over the 80 by Hamilton or Madison. The top-300 count comprises the 51 undisputed Hamilton papers and four of Jay's five, so Jay's side depends on the word list; with them included F (Section 4.3) is 0.574, 0.605, and 0.624.

The hierarchical model's concentrations, estimated from the undisputed papers, are 1,056 (Hamilton) and 1,134 (Madison) under the inherited list, values at which the variance of a paper's counts is roughly twice the multinomial variance. Under that model the pure-author sign is correct for 64, 64, and 62 of the 65 left-out papers under the inherited, closed-class, and top-300 lists, and 0, 2, and 4 of those papers reach a blend factor of 10^2 instead of 3, 7, and 14. Over those fits and the constructed mixtures, with the thirty-word list included (260 and 120 fits), ninety-percent intervals for w contain the truth in 0.68 and 0.64 of cases under word independence and 0.77 and 0.73 under the hierarchical model, against the nominal 0.90. Between undisputed papers the variance of a word's count, in units of its binomial value, has median 1.88 over the seventy inherited words and reaches 10.1 for *his*, pronouns and modals varying most. Giving the ten most variable words a concentration of their own, or every word its own, raises the two coverages to at most 0.81 and 0.78 and leaves No. 55 without a blend preference. In the first variant No. 63 under top-300 gains a blend factor of $10^{2.80}$, the one disputed-paper factor above 10^2 under any overdispersed model.

E Texts, word lists and computation

We describe here the texts, word lists, and computations of Sections 3–6. For the Federalist papers we count words in the Project Gutenberg text, starting after each paper's salutation. Play texts are restricted to verse lines, with early-modern spelling normalized before counting. In the canon survey the profile plays of Marlowe, Kyd, Nashe, and Greene are EEBO-TCP transcriptions. Verse

		$E[w]$		$\log_{10}$ BF(blend : best pure)	
paper	word list	indep.	hier.	indep.	hier.
18	M&W 70 (continuity)	0.0368	0.0156	−1.361	−1.775
	closed class (cc318)	0.0665	0.0258	−0.911	−1.486
	top-300	0.0709	0.0165	−0.880	−1.725
19	M&W 70 (continuity)	0.0579	0.0301	−1.171	−1.495
	closed class (cc318)	0.1604	0.0309	−0.169	−1.485
	top-300	0.2297	0.0374	+0.755	−1.278
20	M&W 70 (continuity)	0.1183	0.0525	−0.655	−1.155
	closed class (cc318)	0.2078	0.0810	+0.243	−0.848
	top-300	0.2579	0.0941	+0.984	−0.382
55	M&W 70 (continuity)	0.5146	0.5154	+2.016	+0.352
	closed class (cc318)	0.3198	0.0749	+2.047	−0.936
	top-300	0.3879	0.1322	+3.966	−0.213

Table 13: Hamilton weight and blend preference in the joint papers and No. 55 under the word-independence model (indep.) and the hierarchical model (hier.): every joint paper stays Madison-dominant with a smaller Hamilton weight, and No. 55's blend preference disappears. $E[w]$: posterior mean of the Hamilton weight; last columns: $\log_{10}$ Bayes factor of the blend against the better pure-author model. Left out in turn under the hierarchical model, Madison's undisputed papers receive $E[w]$ with median 0.095, 0.085, and 0.114 under the three lists (continuity: the inherited list).

makes up 0.779 (*The Tempest*) to 1.000 (*King John*) of the dialogue in the four plays classed "none" and 0.834 to 1.000 in the six with a second hand (*Edward III*'s text, lacking line division, is counted as all verse). It is only 0.421 of the dialogue in *As You Like It*, 0.119 in *Merry Wives*, and 0.276 in *Much Ado*. A second hand writing prose is outside the test, and every power figure of Section 5.2 concerns a second hand writing verse.

In Hebrew and Greek the list categories are lemmas tagged conjunction, preposition, pronoun, or particle (article included, adverbs excluded) and Hebrew pronominal suffixes by person, gender, and number. The tagged texts are the Open Scriptures morphology of the Westminster Leningrad Codex, the ETCBC transcription of the Great Isaiah Scroll (Jacobs et al., 2026), and the SBL Greek New Testament. Where the tagging treats such an item as a suffix, as it does the Aramaic definite article, we count it under the same definition. The canon survey's 365-word list adds early-modern closed-class forms to the 318-word list of Section 4.2. The 142-word *Henry VI* list is the subset of the 318 whose rates are stable across registers. We chose it over the full 318, and over both lists with early-modern forms added, by the criterion of Section 5.3.

We transcribed the tables of Carpenter and Harford-Battersby (1900) twice, independently. After correction against the page scans, the two transcriptions agree on 99.6% of verses (6,466 of the 6,492 compared); we settled the rest from the second printing (Carpenter and Harford, 1902). However, mechanical comparison of the transcription with the 1902 printing's word lists is limited by optical character recognition of pointed Hebrew. All 6,510 verses of the Hexateuch are covered; Deut 23:13–14, absent from both printings, is assigned to D following the accompanying arranged-text volume. Verses that the tables divide between sources carry a single label in Table 7; splitting

or dropping them instead changes the margins by less than 30 nats. Leviticus is a single P range in the block rows (the chapters row subdivides it), and Genesis 14, marked as a P supplement, stays with P throughout.

We computed in exact rational arithmetic every Federalist evidence and posterior of w, every *Henry VI* act and scene value, the twelve Genesis scores underlying Table 7 (four maps, three prior settings), and, in the canon survey, single-profile evidences at every length and blend evidences up to 1,500 list tokens. We did the same for the chapter-level changepoint evidences of up to three processes at each α: nine values per Isaiah manuscript and 36 over the four books of Table 1. All other values are floating point. An independent recomputation for the 41 plays reproduces every classification, with values agreeing to within 4.2×10^{-4}. For the *Henry VI* play rows the floating-point computation agrees with rational arithmetic to within 2×10^{-12} at 4,074 tokens and with 40-digit quadrature to within 3×10^{-11} at 16,055. The floating-point changepoint values are within 10^{-6} of the rational ones in log evidence for Isaiah and within 4.8×10^{-7} for the 36 four-book values. The floating-point entries of Table 7 were computed with two independent programs, which agree to better than 10^{-6}.

For the biblical books we measured detection power on synthetic texts spanning the books' lengths, each with one change of source. The reference strength, Isaiah's measured 1–39 versus 40–66 contrast (total variation 0.192, half the summed absolute difference between the two blocks' function-word rate profiles), lies between the contrast across Esther's two halves (0.110, one source) and that across Zechariah's (0.217, an attested division). Detection in at least 0.8 of replicates needs roughly 2,400, 4,800, and 11,600 function-word tokens at strengths 0.24, 0.19, and 0.14.

As elsewhere (Section 3.4), we wrote down the candidate boundaries for the four books of Table 1 before examining any posterior, though that order rests on the run record alone. The one exception concerns the Pentateuch word list: we compared candidate lists, by within-source prediction only, after seeing a preliminary Genesis computation with the function-word categories; those categories, the list specified in advance, also predicted best.

References

D. Adair, "The authorship of the disputed Federalist papers," The William and Mary Quarterly **1** (1944) 97, and Part II, ibid. 235.

E. M. Airoldi, A. G. Anderson, S. E. Fienberg and K. K. Skinner, "Who wrote Ronald Reagan's radio addresses?," Bayesian Analysis **1** (2006) 289–319.

N. Akiva and M. Koppel, "A generic unsupervised method for decomposing multi-author documents," Journal of the American Society for Information Science and Technology **64**(11) (2013) 2256–2264.

B. Allison and L. Guthrie, "Authorship attribution of e-mail: Comparing classifiers over a new corpus for evaluation," Proc. Sixth International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, European Language Resources Association (2008).

[J. Astruc], *Conjectures sur les mémoires originaux dont il paroit que Moyse s'est servi pour composer le livre de la Genèse*, Bruxelles (1753).

D. Barry and J. A. Hartigan, "Product partition models for change point problems," Ann. Statist. **20** (1992) 260–279.

D. Barry and J. A. Hartigan, "A Bayesian analysis for change point problems," J. Amer. Statist. Assoc. **88** (1993) 309–319.

F. C. Baur, *Paulus, der Apostel Jesu Christi*, Stuttgart (1845).

S. Bozza, V. Cammarota, C.-A. Roten, V. Roten, A. Jover and F. Taroni, "The Bayes factor to discriminate Molière and Corneille as authors of classical French plays," Scientific Reports **15** (2025) 40779.

J. Burrows, "'Delta': a measure of stylistic difference and a guide to likely authorship," Literary and Linguistic Computing **17** (2002) 267.

J. E. Carpenter and G. Harford, *The Composition of the Hexateuch: An Introduction with Select Lists of Words and Phrases*, Longmans, Green, and Co., London (1902).

J. E. Carpenter and G. Harford-Battersby (eds.), *The Hexateuch According to the Revised Version*, 2 volumes, Longmans, Green, and Co., London (1900).

E. K. Chambers, "The disintegration of Shakespeare," Proceedings of the British Academy **11** (1924) 89.

J. Collins, D. Kaufer, P. Vlachos, B. Butler and S. Ishizaki, "Detecting collaborations in text: Comparing the authors' rhetorical language choices in The Federalist Papers," Computers and the Humanities **38** (2004) 15–36.

H. Craig and A. F. Kinney (eds.), *Shakespeare, Computers, and the Mystery of Authorship*, Cambridge University Press, Cambridge (2009).

S. E. De Morgan, *Memoir of Augustus De Morgan*, Longmans, Green (1882), letter of 18 August 1851.

W. M. L. de Wette, *Dissertatio critico-exegetica qua Deuteronomium a prioribus Pentateuchi libris diversum, alius cuiusdam recentioris auctoris opus esse monstratur*, Jena (1805).

I. Dershowitz, N. Akiva, M. Koppel and N. Dershowitz, "Computerized source criticism of biblical texts," Journal of Biblical Literature **134**(2) (2015) 253–271.

J. C. Döderlein, *Esaias ex recensione textus hebraei*, 3rd ed., Altdorf (1789).

B. Duhm, *Das Buch Jesaia*, Handkommentar zum Alten Testament, Vandenhoeck & Ruprecht, Göttingen (1892).

M. Ebrahimpour, T. J. Putniņš, M. J. Berryman, A. Allison, B. W.-H. Ng and D. Abbott, "Automated authorship attribution using advanced signal classification techniques," PLoS ONE **8**(2) (2013) e54998.

P. Edmondson and S. Wells, eds., *Shakespeare beyond Doubt: Evidence, Argument, Controversy*, Cambridge University Press (2013).

J. Eisenstein and R. Barzilay, "Bayesian unsupervised topic segmentation," Proc. 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP 2008) 334–343.

A. Ellegård, *A Statistical Method for Determining Authorship: The Junius Letters, 1769–1772*, Gothenburg Studies in English 13 (1962).

M. Evans, Z. Gilula and I. Guttman, "Latent class analysis of two-way contingency tables by Bayesian methods," Biometrika **76** (1989) 557–563.

S. Faigenbaum-Golovin, A. Kipnis, A. Bühler, E. Piasetzky, T. Römer and I. Finkelstein, "Critical biblical studies via word frequency analysis: unveiling text authorship," arXiv:2410.19883 (2024).

P. Fearnhead, "Exact and efficient Bayesian inference for multiple changepoint problems," Statistics and Computing **16** (2006) 203–213.

M. Font, X. Puig and J. Ginebra, "Bayesian analysis of the heterogeneity of literary style," Revista Colombiana de Estadística **39**(2) (2016) 205–227.

G. Fung, "The disputed Federalist Papers: SVM feature selection via concave minimization," in Proceedings of the 2003 Conference on Diversity in Computing (TAPIA '03), ACM (2003) 42–46.

P. S. Gill and T. B. Swartz, "Stylometric analyses using Dirichlet process mixture models," J. Statist. Plann. Inference **141** (2011) 3665–3674.

P. S. Gill, T. B. Swartz and M. Treschow, "A stylometric analysis of King Alfred's literary works," Journal of Applied Statistics **34**(10) (2007) 1251–1258.

J. Girón, J. Ginebra and A. Riba, "Bayesian analysis of a multinomial sequence and homogeneity of literary style," The American Statistician **59**(1) (2005) 19–30.

M. E. Glickman, J. Brown and R. Song, "(A) Data in the Life: Authorship attribution in Lennon–McCartney songs," Harvard Data Science Review **1**(1) (2019).

T. Grant, *The Idea of Progress in Forensic Authorship Analysis*, Cambridge University Press (2022).

R. A. Hardcastle, "CUSUM: a credible method for the determination of authorship?," Science & Justice **37** (1997) 129.

[S. Hickson], "The shares of Shakspere and Fletcher in The Two Noble Kinsmen," Westminster and Foreign Quarterly Review **48** (1847) 59.

N. L. Hjort, "And Quiet Does Not Flow the Don: Statistical analysis of a quarrel between Nobel prize laureates," Consilience, Centre of Advanced Study, Norwegian Academy of Science and Letters (2007) 134–140, arXiv:2604.19153.

D. I. Holmes and R. S. Forsyth, "The Federalist revisited: New directions in authorship attribution," Literary and Linguistic Computing **10** (1995) 111–127.

H. J. Holtzmann, *Die synoptischen Evangelien: ihr Ursprung und geschichtlicher Charakter*, Leipzig: Engelmann (1863).

C. Hoy, "The shares of Fletcher and his collaborators in the Beaumont and Fletcher canon (I)," Studies in Bibliography **8** (1956) 129–146.

H. Hupfeld, *Die Quellen der Genesis und die Art ihrer Zusammensetzung*, Berlin: Wiegandt und Grieben (1853).

S. Ishihara, "Weight of authorship evidence with multiple categories of stylometric features: A multinomial-based discrete model," Science & Justice **63**(2) (2023) 181–199.

M. P. Jackson, *Defining Shakespeare: Pericles as Test Case*, Oxford University Press (2003).

J. Jacobs, M. Naaijer and D. Roorda, *The Dead Sea Scrolls in Text-Fabric*, dataset version 2.0 (2026), ETCBC, Vrije Universiteit Amsterdam, `github.com/ETCBC/dss`, Zenodo concept DOI `10.5281/zenodo.2652849`; based on the electronic transcriptions and morphological tagging of M. G. Abegg, Jr.

S. W. Jeong and V. Ročková, "From small to large language models: Revisiting the Federalist Papers," arXiv:2503.01869.

M. L. Jockers and D. M. Witten, "A comparative study of machine learning methods for authorship attribution," Literary and Linguistic Computing **25**(2) (2010) 215–223.

P. Juola, "The Rowling case: A proposed standard analytic protocol for authorship questions," Digital Scholarship in the Humanities **30** (2015) i100.

J. Kim, "Marginal likelihoods for finite-support Huber contamination," arXiv:2605.26723 (2026).

M. Koppel, J. Schler and S. Argamon, "Computational methods in authorship attribution," Journal of the American Society for Information Science and Technology **60** (2009) 9.

M. Koppel, N. Akiva, I. Dershowitz and N. Dershowitz, "Unsupervised decomposition of a document into authorial components," Proc. 49th Annual Meeting of the ACL: Human Language Technologies (2011) 1356–1364.

E. Y. Kutscher, *The Language and Linguistic Background of the Isaiah Scroll (1 Q Isa^a)*, Studies on the Texts of the Desert of Judah 6, Brill, Leiden (1974).

S. Lin, B. Sturmfels and Z. Xu, "Marginal likelihood integrals for mixtures of independence models," J. Mach. Learn. Res. **10** (2009) 1611–1631, [arXiv:0805.3602].

J. S. Liu and C. E. Lawrence, "Bayesian inference on biopolymer models," Bioinformatics **15**(1) (1999) 38–52.

H. Love, *Attributing Authorship: An Introduction*, Cambridge University Press (2002).

W. Lutosławski, *The Origin and Growth of Plato's Logic, with an Account of Plato's Style and of the Chronology of his Writings*, Longmans, Green (1897).

R. E. Madsen, D. Kauchak and C. Elkan, "Modeling word burstiness using the Dirichlet distribution," Proc. 22nd International Conference on Machine Learning (ICML 2005), ACM (2005) 545–552.

E. Malone, *A Dissertation on the Three Parts of King Henry VI*, London (1787).

T. C. Mendenhall, "The characteristic curves of composition," Science **ns-9** (1887) 237.

T. C. Mendenhall, "A mechanical solution of a literary problem," Popular Science Monthly **60** (1901) 97.

G. Meshcheryakov, "Exact computation of posterior distribution of mixture weights in hierarchical Bayesian models," arXiv:2607.05692.

J. E. Mosimann, "On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions," Biometrika **49** (1962) 65.

F. Mosteller and D. L. Wallace, "Inference in an authorship problem," J. Amer. Statist. Assoc. **58** (1963) 275–309.

F. Mosteller and D. L. Wallace, *Inference and Disputed Authorship: The Federalist*, Addison–Wesley, Reading MA (1964).

P. Plecháč, "Relative contributions of Shakespeare and Fletcher in Henry VIII: An analysis based on most frequent words and most frequent rhythmic patterns," Digital Scholarship in the Humanities **36** (2021) 430–438, arXiv:1911.05652.

M. Popović, M. A. Dhali and L. Schomaker, "Artificial intelligence based writer identification generates new evidence for the unknown scribes of the Dead Sea Scrolls exemplified by the Great Isaiah Scroll (1QIsa$^{\mathrm{a}}$)," PLOS ONE **16** (2021) e0249769.

X. Puig, M. Font and J. Ginebra, "Classification of literary style that takes order into consideration," Journal of Quantitative Linguistics **22**(3) (2015) 177–201.

X. Puig, M. Font and J. Ginebra, "A unified approach to authorship attribution and verification," The American Statistician **70**(3) (2016) 232–242.

Y. T. Radday, "Isaiah and the computer: A preliminary report," Computers and the Humanities **5**(2) (1970) 65–73.

Y. T. Radday, *The Unity of Isaiah in the Light of Statistical Linguistics*, Gerstenberg, Hildesheim (1973).

Y. T. Radday and H. Shore, *Genesis: An Authorship Study in Computer-Assisted Statistical Linguistics*, Analecta Biblica 103, Biblical Institute Press, Rome (1985).

M. Rosen-Zvi, T. Griffiths, M. Steyvers and P. Smyth, "The author-topic model for authors and documents," Proc. 20th Conference on Uncertainty in Artificial Intelligence (2004) 487–494.

G. J. Ross, "Tracking the evolution of literary style via Dirichlet–multinomial change point regression," J. R. Statist. Soc. A **183**(1) (2020) 149–167.

J. Rudman, "The twelve disputed 'Federalist' papers: A case for collaboration," Digital Humanities 2012 (Hamburg, 2012), conference abstracts.

J. Savoy, "Is Luke the author of a Gospel and the Acts of the Apostles?," arXiv:2609.17762 (2026).

K. Schmid, "The Neo-Documentarian Manifesto: A critical reading," Journal of Biblical Literature **140** (2021) 461.

S. Schoenbaum, *Internal Evidence and Elizabethan Dramatic Authorship*, Northwestern University Press (1966).

M. D. Schwartz, "Exact Bayesian evidence for mixture models, and the millions of observations model selection needs," companion paper (2026), `https://bootloops.ai/mixtures.pdf`.

S. Segarra, M. Eisen, G. Egan and A. Ribeiro, "Attributing the authorship of the Henry VI plays by word adjacency," Shakespeare Quarterly **67** (2016) 232–256.

J. S. Semler, annotated paraphrase of Second Corinthians, Halle (1776); partition attribution per J. H. Ropes, Encyclopædia Britannica, 11th ed.

Y. Seroussi, I. Zukerman and F. Bohnert, "Authorship attribution with topic models," Computational Linguistics **40** (2014) 269–310.

J. Spedding, "Who wrote Shakspere's Henry VIII?," Gentleman's Magazine, n.s. **34** (1850) 115.

E. Stamatatos, "A survey of modern authorship attribution methods," Journal of the American Society for Information Science and Technology **60** (2009) 538.

B. H. Streeter, *The Four Gospels: A Study of Origins*, Macmillan (1924).

G. Taylor and G. Egan (eds.), *The New Oxford Shakespeare: Authorship Companion*, Oxford University Press, Oxford (2017).

E. Ulrich and P. W. Flint, with a contribution by M. G. Abegg, Jr., *Qumran Cave 1.II: The Isaiah Scrolls*, 2 parts, Discoveries in the Judaean Desert XXXII, Clarendon Press, Oxford (2010).

M. Utiyama and H. Isahara, "A statistical model for domain-independent text segmentation," Proc. 39th Annual Meeting of the ACL (2001) 499–506.

B. Vickers, *Shakespeare, Co-Author: A Historical Study of Five Collaborative Plays*, Oxford University Press, Oxford (2002).

B. Vickers, "Shakespeare and authorship studies in the twenty-first century," Shakespeare Quarterly **62** (2011) 106–142.

J. Wellhausen, *Prolegomena to the History of Israel*, translated by J. S. Black and A. Menzies, Adam & Charles Black, Edinburgh (1885).

D. Wickmann, *Eine mathematisch-statistische Methode zur Untersuchung der Verfasserfrage literarischer Texte*, Westdeutscher Verlag, Köln und Opladen (1968).

G. Yoffe, A. Bühler, N. Dershowitz, I. Finkelstein, E. Piasetzky, T. Römer and B. Sober, "A statistical exploration of text partition into constituents: The case of the Priestly source in the books of Genesis and Exodus," arXiv:2305.02170 (2023).

G. U. Yule, "On sentence-length as a statistical characteristic of style in prose, with application to two cases of disputed authorship," Biometrika **30** (1939) 363.

G. U. Yule, *The Statistical Study of Literary Vocabulary*, Cambridge University Press (1944).

E. Zangerle, M. Tschuggnall, G. Specht, B. Stein and M. Potthast, "Overview of the style change detection task at PAN 2019," CLEF 2019 Working Notes, CEUR Workshop Proceedings **2380** (2019).