Research Article

A Statistical Investigation of the Synoptic Problem

David Gentile[1]

1. Independent researcher

This study presents a statistical investigation of the literary relationships among the three synoptic gospels — Matthew, Mark, and Luke — using vocabulary frequency data drawn from the Hoffmann, Hieke, and Bauer Synoptic Concordance. Word occurrence counts across nineteen synoptic categories are modeled using a Poisson distribution, and pairwise relationships between categories are tested using maximum likelihood estimation and the likelihood ratio test, with a correction for the overdispersion characteristic of word-frequency data. The analysis identifies statistically significant vocabulary affinities between specific synoptic categories, yielding evidence for Markan priority, for Luke's use of Matthew, and for a sayings source underlying the double tradition. The results are most consistent with the Three Source Hypothesis: Mark composed first; Matthew drew on Mark and a sayings source; Luke drew on Mark, the sayings source, and Matthew. The same framework accommodates, without establishing, more speculative refinements — a proto-Mark and a proto-Luke — that are developed on literary grounds in the discussion and reserved for fuller treatment elsewhere. An extended example from Luke 14:1–6 illustrates how an alliterative Aramaic saying, preserved in Greek translation, illustrates Luke's use of all three sources.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

1. Introduction

The synoptic problem — the question of the literary relationship among the gospels of Matthew, Mark, and Luke — has occupied scholars for more than two centuries. The three gospels share extensive verbal agreement, parallel narrative sequences, and common sayings material, yet they also differ in ways that resist any simple explanation. For a comprehensive overview of the problem, the proposed solutions, and the relevant primary and secondary literature, readers are directed to [1].

The dominant solution in modern scholarship is the Two Source Hypothesis (2SH), which holds that Mark was composed first and that both Matthew and Luke independently drew on Mark and on a lost sayings document conventionally designated Q. [2][3] The Q hypothesis has been developed most fully by Kloppenborg [4], whose stratigraphic analysis of the document's layers has been especially influential. The 2SH accounts for the broad patterns of agreement and disagreement but leaves several features of the tradition unexplained — most notably the "minor agreements," passages where Matthew and Luke agree against Mark in ways difficult to attribute to independent editorial coincidence, and the Mark/Q overlaps, which show the theory to be at the least incomplete.

Alternative hypotheses have been proposed. The Farrer Hypothesis (FH) dispenses with Q entirely, holding that Luke used Matthew directly in addition to Mark [5]. The Griesbach Hypothesis (GH) [6], revived in the twentieth century, holds that Matthew was composed first, that Luke used Matthew, and that Mark is an abbreviation of both. A further option, the Three Source Hypothesis (3SH), retains both Q-like sayings material and direct dependence, holding that Luke drew on Mark, on a sayings source, and on Matthew [7][8][9][10]; it is this solution that the present study supports. Each hypothesis has its defenders, and the debate has proceeded largely on the basis of qualitative literary arguments — assessments of editorial tendency, plausibility of borrowing direction, and the relative primitiveness of particular readings.

The present study brings a different kind of evidence to bear. Rather than analyzing individual passages, it examines the statistical behavior of vocabulary across the entire synoptic corpus. The underlying logic is straightforward: if two categories of synoptic material were originally composed by the same author, we would expect their vocabulary profiles to be more similar to each other than to the synoptic corpus as a whole. Shared subject matter can also raise the affinity between two categories, a confound addressed in the analysis; but where the pattern of affinities is directional — where one author's unique category predicts a shared category and another's does not — the most economical explanation is a relationship of composition and source. By testing all pairwise relationships between the nineteen synoptic categories defined in the HHB Synoptic Concordance [11], the study assigns quantitative probabilities to these relationships and identifies which are too strong to be explained by chance. As will be seen, the results establish the priority of Mark beyond reasonable doubt, against the Griesbach Hypothesis; they tell strongly against the independence of Matthew and Luke that the Two Source Hypothesis requires; and they support a sayings source, against the Farrer Hypothesis.

A good starting point is to show, diagrammatically, what the study is trying to do. In the example below we are trying to determine which author wrote first and which was the editor. Two cases are presented, one where author X wrote first and one where author Y wrote first. The words in bold are by author X and the words in italic are by author Y. Note that it is the words they share in common whose authorship is in question. We can build a vocabulary profile of each author from the words we are sure they wrote, and then see which author's style corresponds with the words of undetermined authorship.

Case 1 - Author X wrote first													
The	**quick**	**brown**	**fox**	**jumped**	**over**	**the**	**lazy**	**dogs**					
The	*speedy*	**brown**	*animal*	**jumped**					*a*	*fence*	*and*	*it*	*escaped*
Case 2 - Author Y wrote first													
The	**quick**	*brown*	**fox**	*jumped*	**over**	**the**	**lazy**	**dogs**					
The	*speedy*	*brown*	*animal*	*jumped*					*a*	*fence*	*and*	*it*	*escaped*

The study proceeds as follows. The Data section describes the HHB Synoptic Concordance and the structure of the dataset. The Synoptic Categories subsection explains the nineteen category codes used throughout. The Analysis and Mathematics sections describe the statistical methods, including the correction for overdispersion in the count data. The Results section presents all statistically significant findings. The Discussion section interprets the results in terms of the competing hypotheses and considers more speculative refinements — a proto-Mark and a proto-Luke — argued there on literary rather than statistical grounds. Conclusions follow.

Supporting materials are provided as supplements: full definitions of the nineteen synoptic categories, together with the readings each competing hypothesis assigns them (Supplementary Appendix A); the complete enumeration of the sixty-six hypothesis-testing comparisons to which the multiple-comparison corrections are applied (Supplementary Appendix B); the Excel spreadsheet holding the raw frequency data and the computational machinery for the pairwise analysis, with a guide to its organization; and an extended worked example from Luke 14:1–6 that illustrates the Three Source Hypothesis at the level of a single pericope, tracing in Luke's text the convergence of Markan narrative, an Aramaic-derived saying from the sayings source, and Matthew's editorial combination of the two.

2. The Data

The dataset for this study is drawn from the Synoptic Concordance compiled by [11], a systematic four-volume tabulation of vocabulary occurrences across the three synoptic gospels organized by synoptic

category. The concordance represents the most comprehensive lexical resource of its kind, and its category structure — which assigns every word occurrence to one of nineteen synoptic categories based on patterns of agreement and divergence among the three gospels — makes it uniquely suited to the kind of statistical analysis undertaken here.

Of the vocabulary items tabulated in the HHB concordance, 807 are covered with complete frequency data across all synoptic categories; the rarer items receive less complete tabular treatment and are excluded from the present analysis. This restriction is not merely a constraint of the source but an advantage for the statistics: the excluded items are low-frequency words whose sparse counts would yield unstable frequency estimates and inflate the dispersion of the data, while the 807 retained items are well enough attested to support reliable estimation. The dataset used here therefore consists of these 807 fully tabulated vocabulary items, each treated as a row in the underlying spreadsheet, with nineteen columns corresponding to the nineteen synoptic categories. Each cell records the absolute frequency of a given vocabulary item in a given category — that is, the raw count of occurrences.

To make the structure of the dataset concrete: imagine working through a synopsis of the three gospels with nineteen colored pens, one per synoptic category. Counting, for each vocabulary item, how many instances fall under each color would reproduce the raw data table exactly.

The Synoptic Categories. The HHB concordance organizes every word occurrence in the synoptic gospels into one of nineteen categories, determined by the pattern of agreement and divergence among the three gospels at that point in the tradition. Each category is designated by a three-digit code in which the digits represent Matthew, Mark, and Luke respectively. A "2" indicates that the gospel contains the word in question, in significant verbal agreement with the parallel; a "1" indicates that the gospel has a corresponding parallel passage but without the word; and a "0" indicates that the gospel has no corresponding passage at all.

The system is best grasped through a small set of illustrative cases before the full taxonomy is consulted (see Supplementary Appendix A).

222 designates the triple tradition with full verbal agreement — passages where all three gospels share the word. On the dominant Two Source Hypothesis, this material represents Markan vocabulary copied independently by both Matthew and Luke; on the Griesbach Hypothesis, Matthean vocabulary copied by Luke and then Mark. The competing hypotheses read categories differently in this way.

221 designates passages where Matthew and Mark share the word and Luke has the parallel passage but not the word. The Two Source Hypothesis assigns this to Mark, copied by Matthew but not Luke; the Griesbach Hypothesis, to Matthew, copied by Mark but not Luke.

122 designates passages where Mark and Luke share the word and Matthew has the parallel passage but not the word — on the Two Source Hypothesis, Markan material copied by Luke but not Matthew.

212 designates passages where Matthew and Luke share the word and Mark has the parallel passage but not the word — the so-called "minor agreements", including the major minor agreements or what the 2SH calls the Mark/Q overlaps. These are among the most theoretically contested data in synoptic scholarship, since they are difficult to account for if, as the Two Source Hypothesis holds, Matthew and Luke worked independently of one another.

200, 020, and 002 designate material unique to Matthew, Mark, and Luke respectively — the Sondergut, or "special material," categories, in which no cross-gospel agreement is at issue.

Familiarity with the category system is presupposed throughout the Results and Discussion sections.

3. The Analysis

The statistical method used in this study tests whether the vocabulary profile of one synoptic category can serve as a significant predictor of the vocabulary profile of another. The mathematical details are given in the Mathematics section; what follows is a conceptual account of the logic.

Definitions. Because the word "word" is ambiguous in this context — referring variously to a lexical item, a particular inflected form, or a single occurrence in the text — the term *vocabulary item* is used throughout to refer to the 807 lexical entries that constitute the rows of the dataset. A single instance of a vocabulary item appearing in the text is called an *occurrence*. The *absolute frequency* of a vocabulary item in a given category is the raw count of its occurrences there. The *relative frequency* is that count divided by the total number of occurrences of all vocabulary items in the category — a normalized measure that allows meaningful comparison across categories of different sizes. If category 222 contains 2,000 total occurrences and πνεῦμα appears 10 times within it, its relative frequency in that category is 10/2,000, or 1/200.

The core question. The central question the study asks of each pair of categories is: does knowing the relative frequency of vocabulary items in category A improve our ability to predict their relative frequency in category B, beyond what the overall synoptic vocabulary distribution already tells us? If it

does — consistently, across hundreds of vocabulary items, to a degree that cannot plausibly be attributed to chance — that constitutes evidence of a special relationship between the two categories. The most natural explanation for such a relationship is shared authorship — the two categories were composed by the same hand — though shared subject matter can also raise affinity, a confound the analysis addresses by attending to the *direction* of prediction (which category predicts which) rather than affinity alone.

The baseline against which each category is tested is the relative frequency of each vocabulary item across all synoptic categories combined, minus the test category itself. Subtracting the test category from the baseline is a refinement introduced in the present version of the study; it prevents the category being predicted from artificially inflating its own baseline and produces a cleaner test of predictor influence.

The method. For each of the 171 possible pairwise comparisons among the nineteen categories, the study asks whether adding a second category as a predictor significantly improves the fit of a Poisson model to the observed frequency data in the test category. The Poisson distribution is appropriate here because the data consist of counts of non-negative integers — word occurrences cannot be fractional or negative — and it plays the same role that the normal distribution plays in standard regression, or the Bernoulli distribution in logistic regression. The improvement in fit achieved by adding the predictor category is evaluated using the likelihood ratio test, which yields a probability that the observed improvement could have arisen by chance alone. Pairs of categories for which this probability falls below the significance threshold are identified as having a statistically meaningful relationship.

4. The Mathematics

The Poisson distribution. The statistical model underlying this study treats word occurrences as count data governed by a Poisson distribution. [12][13] The Poisson distribution is analogous in many respects to the normal distribution: just as the normal distribution is characterized by its mean and standard deviation, the Poisson distribution is characterized by a single parameter, here denoted γ (gamma), which identifies the expected rate and around which the probability of other outcomes decreases with distance. The crucial difference is that the Poisson distribution is defined only over non-negative integers, making it the natural choice for modeling count data such as word occurrences, where fractional or negative values are impossible.

The parameter γ can be understood as the expected rate of occurrence. If a given vocabulary item appears on average three times per thousand words of ancient Greek, then $\gamma = 3$ for that item in a thousand-word sample, and the Poisson distribution gives the probability of observing any particular count — two

occurrences, five, zero — in a new sample of the same length. the single most probable count is the integer nearest γ; departures in either direction become progressively less probable.

Maximum likelihood estimation. The maximum likelihood procedure [13] inverts this logic. Rather than using a known γ to predict the probability of observed counts, it asks: given a set of observed counts, what value of γ makes those observations most probable? If a vocabulary item is observed 10, 12, and 8 times across three comparable samples, a γ of 10 assigns those observations far higher probability than a γ of 2 would. Maximum likelihood estimation finds the γ that maximizes this probability across all observations simultaneously.

Application to the synoptic data. In the present study, the initial γ for each vocabulary item in each test category is set by that item's relative frequency across all synoptic categories combined, excluding the test category, scaled by the total word count of the test category. This baseline γ encodes the null hypothesis: that the test category draws on the same vocabulary distribution as the synoptic corpus at large, with no special affinity for any particular predictor category.

The study then asks whether introducing a predictor category improves on this null model. A single weighting parameter β (beta) is added. The adjusted γ for each vocabulary item is computed as a weighted combination of its relative frequency in the overall corpus and its relative frequency in the predictor category:

$$\gamma_{adjusted} = (1 - \beta) \cdot \gamma_{baseline} + \beta \cdot \gamma_{predictor}$$

When β = 0, the predictor category contributes nothing and the model reduces to the null. When β > 0, the predictor category pulls the expected frequencies toward its own profile. The value of β that maximizes the joint probability of all 807 observed counts in the test category simultaneously is found using Excel's Solver function.

The likelihood ratio test. The improvement in fit achieved by introducing β is evaluated using the likelihood ratio test. [13] The test statistic is twice the difference in log-likelihood between the fitted model and the null model, which under standard conditions follows a chi-squared distribution with one degree of freedom. This yields a p-value: the probability that an improvement of the observed magnitude would arise by chance if the predictor category in fact carried no information about the test category. It is this p-value that is recorded in the results matrix and reported in the Results section.

To summarize, the test for each ordered pair of categories draws on three inputs:

1. The relative frequency of 807 vocabulary items across all synoptic categories combined, excluding the test category — the baseline.
2. The absolute frequency of those 807 items in the test category — what is to be predicted.
3. The relative frequency of those 807 items in the predictor category — the candidate source of additional information.

The question is whether source 3, combined with source 1, predicts source 2 significantly better than source 1 alone. For comprehensive treatments of maximum likelihood estimation and the likelihood ratio test as applied to count data, see [13] and [14].

Overdispersion and the dispersion correction. The Poisson model assumes that the variance of each count equals its mean. Word-frequency data violate this assumption: vocabulary clusters by topic and discourse type, so once a word appears it tends to recur nearby rather than arriving at a constant rate. A passage of direct address is dense with second-person pronouns; a healing narrative is dense with a different vocabulary again. This "burstiness" makes the observed counts scatter more widely than Poisson predicts — a condition known as overdispersion — and an uncorrected likelihood ratio test will, in consequence, return p-values that are too small, overstating significance. [14]

The magnitude of the effect can be estimated directly. For a fitted comparison, the dispersion factor φ is the mean of the squared standardized residuals — Σ(observed − expected)² / expected, divided by the degrees of freedom. Because φ measures the dispersion of the count data and not the strength of any relationship, it can be estimated from whichever comparisons have the densest counts, independent of whether those comparisons belong to the hypothesis-testing family. Sparse cells would distort our estimate so we excluded cells with expected values less than one – reducing the degrees of freedom. Thus:

$$\hat{\varphi} = [\Sigma_i (y_i - \hat{\mu}_i)^2 / \hat{\mu}_i] / (n_{retained\ cells} - 1)$$

Estimated from three well-populated test categories, where expected counts are large enough that the statistic is not distorted by sparse cells, φ falls in a narrow band: 1.88 for predictions into the triple-agreement category 222 and 1.76 for the Mark/Luke agreement category 122 — both within the test family — with a prediction into the well-populated Sondergut-Luke category 002 giving 1.80 in confirmation. The representative value is φ ≈ 1.8. The estimate is stable across comparisons and is not driven by any single item; its largest single contributor, in the 002 estimate, is the second-person

pronoun ὑμῖν, whose concentration in discourse material is exactly the kind of topical clumping the correction is meant to absorb.

The correction is applied by dividing each likelihood ratio statistic by φ before referring it to the chi-squared distribution — equivalently, by inflating each reported p-value. Because the same factor is applied to every test, the correction cannot reverse the *direction* of any relationship or alter the *relative* standing of competing predictors; it uniformly discounts the strength of the evidence to reflect the burstiness of the data. The corrected significance of every reported relationship is given in Table 2.

5. Results

Significance threshold. Not every pairwise comparison among the nineteen categories tests a contested hypothesis. A comparison between two categories distinctive to the same gospel, for instance, asks whether an author's vocabulary predicts his own — a relationship expected on every synoptic hypothesis and therefore uninformative. The comparisons that bear on the competing hypotheses are those that predict one of the seven cross-gospel agreement categories (222, 221, 212, 122, 220, 202, and 022), where the hypotheses make differing claims about authorship and direction of borrowing. Restricting the family to these — every candidate predictor of the triple-agreement category 222 (all author specific categories are relevant of course, but also, even categories shared by two authors are relevant if they predict material shared by all three) together with the hypothesis-relevant predictors of the six other agreement categories — yields 66 hypothesis-testing comparisons; the full enumeration is given in supplementary Appendix B. This is the family to which the multiple-comparison correction is applied. The exclusion rule is content-based and fixed independently of the results: a comparison is in the family if it could discriminate between hypotheses, not if it happened to reach significance. Against this family the Bonferroni threshold is $.05 / 66 \approx 7.6 \times 10^{-4}$. The correction is applied on top of the dispersion correction ($\varphi = 1.8$) described above, so the reported significances remain doubly conservative. As a robustness floor, we note that even against the full set of 171 possible category pairs — the most severe family one could reasonably define — the seven strongest relationships still clear Bonferroni, so the central conclusions do not depend on the narrower family.

Reported relationships. Several categories of results have been excluded from the table below. Relationships between categories whose shared authorship is uncontested — such as one Markan category predicting another — are omitted as uninformative. Uninteresting reciprocal relationships are likewise suppressed. If we know a predictor category predicts a test category, the test category also

predicting the predictor category is of little interest. Note that the evidential asymmetry exploited in this study is of a different kind: not whether one category predicts another versus the reverse, but which of several candidate predictors successfully predicts a single shared target. i.e does 211 => 221 or does 121 => 221? The relationship between categories 200 and 202 is excluded on methodological grounds: the HHB concordance distributes Matthean doublets across these two categories putting one half of each pair in each category. Obviously, the same sentence will look like itself, rendering the result a possible artifact of the concordance's editorial decisions rather than a genuine literary signal. Taking note of this difficulty is one of the improvements from an older version of this study. This is a deliberately conservative choice. The eightteen relationships that survive these filters are reported in Table 1.

Predictor	Test	p-value
221	222	6×10^{-9}
122	222	9×10^{-5}
022	222	5×10^{-5}
220	222	3×10^{-12}
211	212	9×10^{-7}
210	212	7×10^{-9}
121	221	2×10^{-18}
021	221	2×10^{-6}
020	221	5×10^{-5}
121	122	3×10^{-4}
120	122	1×10^{-4}
021	022	7×10^{-5}
120	220	3×10^{-4}
020	220	2×10^{-4}
201	202	5×10^{-10}
102	202	2×10^{-8}
122	002	1×10^{-12}
120	221	6×10^{-4}

Table 1. Statistically significant pairwise relationships ($p < 7.6 \times 10^{-4}$).

A false-discovery-rate alternative. The Bonferroni threshold controls the probability of even a single false positive across the family — appropriate where any false claim is costly, but unduly severe when most of the comparisons in the family are expected to be genuine. The Benjamini–Hochberg false-discovery-rate (FDR) procedure instead controls the expected proportion of false positives among the

relationships declared significant, and is the more appropriate correction here. Applied to the exact dispersion-corrected p-values across the 66-comparison family at q = 0.05, all seventeen of the reported within-family relationships clear; under the stricter Bonferroni threshold, eight (of the 66-family) clear. The result is a graduated picture — the strongest relationships survive even the harshest correction against the widest family, while the full set of seventeen survives the correction appropriate to a study of this design.

As a further check that the conclusions do not depend on the precise dispersion estimate, the correction was recomputed at a deliberately conservative $\varphi = 2.5$, well above the estimated value.

(In a subsequent check, we estimated φ using the study's three most important results 121 → 221, 201 → 202, 102 → 202 and we computed a value of $\varphi = 2.17$.That this estimate is drawn from the principal comparisons themselves is not a circularity but the standard quasi-Poisson procedure, in which the dispersion factor is estimated from a fitted model's own Pearson residuals. At $\varphi = 2.17$, thirteen of the seventeen within-family relationships clear the false-discovery-rate threshold; the four that do are redundant corroborators of Markan priority whose anchor results clear even the strict Bonferroni threshold. No principal conclusion rests on a relationship that fails at this inflated dispersion factor).

Even at the inflated value of $\varphi = 2.5$, six relationships clear the strict Bonferroni threshold without recourse to the FDR procedure — the "robust core" marked in Table 2 — and these span every principal finding: Markan priority, the Markan origin of the triple-tradition agreements, the sayings source behind the double tradition, and the minor agreements as Matthean. The central conclusions are thus robust to substantial misestimation of the dispersion factor.

One reported relationship, 122 → 002, falls outside this family: it predicts a Sondergut category rather than an agreement category, and so tests no competing hypothesis. It is reported as an incidental, exploratory finding rather than a confirmatory one — though it is strong enough (corrected $p = 2.0 \times 10^{-6}$) to clear even the most conservative correction against the full set of pairwise comparisons.

Predictor → Test	raw p	corrected p (φ = 2.17)	significance (Bonferroni / FDR)
121 → 221	2×10^{-18}	7.2×10^{-9}	yes (robust core at φ = 2.5)
220 → 222	3×10^{-12}	2.0×10^{-6}	yes (robust core at φ = 2.5)
201 → 202	5×10^{-10}	2.0×10^{-5}	yes (robust core at φ = 2.5)
221 → 222	6×10^{-9}	8.0×10^{-5}	yes (robust core at φ = 2.5)
210 → 212	7×10^{-9}	8.0×10^{-5}	yes (robust core at φ = 2.5)
102 → 202	2×10^{-8}	2.0×10^{-4}	yes (robust core at φ = 2.5)
211 → 212	9×10^{-7}	8.0×10^{-4}	yes at φ = 1.8, only under FDR at φ = 2.17
021 → 221	2×10^{-6}	1.2×10^{-3}	yes at φ = 1.8, only under FDR at φ = 2.17
020 → 221	5×10^{-5}	5.9×10^{-3}	significant only under FDR
022 → 222	5×10^{-5}	6.0×10^{-3}	significant only under FDR
021 → 022	7×10^{-5}	7.0×10^{-3}	significant only under FDR
122 → 222	9×10^{-5}	7.9×10^{-3}	significant only under FDR
120 → 122	1×10^{-4}	8.6×10^{-3}	significant only under FDR
020 → 220	2×10^{-4}	1.2×10^{-2}	fails FDR at φ = 2.17
121 → 122	3×10^{-4}	1.5×10^{-2}	fails FDR at φ = 2.17
120 → 220	3×10^{-4}	1.3×10^{-2}	fails FDR at φ = 2.17
120 → 221	6×10^{-4}	1.9×10^{-2}	fails FDR at φ = 2.17
122 → 002	1×10^{-12}	2.0×10^{-6}	exploratory (outside family)

Table 2. Reported relationships evaluated against the 66-comparison hypothesis-testing family. The corrected-*p* column is computed at the dispersion factor of φ = 2.17, estimated directly from the study's three principal comparisons by the standard quasi-Poisson procedure; the outcomes at two other values are encoded in the significance column rather than shown as separate columns. Those values are φ = 1.8, from our original estimate, and φ = 2.5, a deliberately conservative stress value well above either estimate. The Bonferroni threshold is .05 / 66 ≈ 7.6×10^{-4}; the FDR procedure is Benjamini–Hochberg at q = 0.05 across the family. All seventeen within-family relationships clear FDR at φ = 1.8; the parenthetical annotations record

doi.org/10.32388/RIBDZ7.3

how each relationship's standing changes as φ is raised. The four significance labels are: "yes (robust core at φ = 2.5)" — clears strict Bonferroni even at φ = 2.5; these six span every principal finding (Markan priority, the Markan origin of the triple-tradition agreements, the sayings source, and the minor agreements as Matthean). "yes at φ = 1.8, only under FDR at φ = 2.17" — clears Bonferroni at φ = 1.8 but, at φ = 2.17, clears only the FDR threshold. "significant only under FDR" — clears FDR but not Bonferroni at φ = 1.8, and continues to clear FDR at φ = 2.17. "fails FDR at φ = 2.17" — clears FDR at φ = 1.8 but not at φ = 2.17; all four are redundant corroborators of Markan priority whose anchor results survive at the higher φ. The corrected value for 121 → 221 uses the approximation $p^{\wedge}(1/\varphi)$, its raw p being too small for stable inversion of the χ^2 quantile; all others use the exact φ-scaled statistic. The relationship 122 → 002 lies outside the family (it predicts a Sondergut category) and is reported as exploratory.

Markan priority over Matthew. Six results (shown below) establish that Mark's distinctive categories predict the material Mark shares with Matthew — most decisively 121 → 221, a relationship between two large triple-tradition categories and significant at 7.2×10^{-9} after correction — while Matthew's distinctive categories do not — showing that Mark's vocabulary profile is antecedent to Matthew's:

- 121 → 221 (corrected $p = 7.2 \times 10^{-9}$)
- 021 → 221 (corrected $p = 1.2 \times 10^{-3}$) significant at φ = 1.8, only under FDR at φ = 2.17
- 020 → 221 (corrected $p = 5.9 \times 10^{-3}$) significant only under FDR
- 120 → 221 (corrected $p = 1.9 \times 10^{-2}$) fails FDR at φ = 2.17
- 020 → 220 (corrected $p = 1.2 \times 10^{-2}$) fails FDR at φ = 2.17
- 120 → 220 (corrected $p = 1.3 \times 10^{-2}$) fails FDR at φ = 2.17

The Mark/Matthew agreement categories 221 and 220 are predicted by Mark's distinctive categories (121, 021, 020, and 120 predict 221; 020 and 120 predict 220) but by none of Matthew's distinctive categories (211, 201, 210, 200). The shared material carries Mark's vocabulary fingerprint, not Matthew's. Were Matthew the source — as the Griesbach Hypothesis holds — 211 should, for example, predict 221 and the Markan predictors should not; the data show the reverse.

Markan priority over Luke. Three results bear on the relationship between Markan and Lukan categories:

- 021 → 022 (corrected $p = 7.0 \times 10^{-3}$) significant only under FDR

- 120 → 122 (corrected p = 8.6×10^{-3}) significant only under FDR
- 121 → 122 (corrected p = 1.5×10^{-2}) fails FDR at $\varphi = 2.17$

These are the results most affected by the dispersion correction: under $\varphi = 2.17$ each falls to conventional significance ($p \approx .007–.015$), below the strict Bonferroni threshold but, as it turns out, two fall comfortably within the false-discovery-rate threshold. So priority over Luke does receive direct support under the correction appropriate to a study of this design. It does not, however, rest on these direct relationships alone, for it also follows transitively from two findings that survive even the strict correction with room to spare: Markan priority over Matthew (121 → 221, corrected $p \approx 10^{-8}$) and Luke's use of Matthew (the minor-agreement results below). If Mark is prior to Matthew and Luke drew on Matthew, Mark is necessarily prior to Luke. The directional asymmetry points the same way and is unaffected by φ: distinctively Markan categories predict the Mark/Luke shared material more strongly than distinctively Lukan categories do. That the direct Mark→Luke signal is weaker than the Mark→Matthew signal is itself expected, and is taken up below in connection with proto-Mark.

The triple tradition, triple agreements. Four results indicate that the two-gospel agreement categories that include Mark predict the triple-agreement category 222 — 221 and 220 (Mark with Matthew) and 122 and 022 (Mark with Luke). In each the shared vocabulary is one Mark carries, and its predicting the triple agreement indicates that the word common to all three gospels entered the tradition through Mark.

- 221 → 222 (corrected p = 8.0×10^{-5})
- 220 → 222 (corrected p = 2.0×10^{-6})
- 122 → 222 (corrected p = 7.9×10^{-3}) significant only under FDR
- 022 → 222 (corrected p = 6.0×10^{-3}) significant only under FDR

The double tradition. Two results bear on the origin of the double-tradition material:

- 201 → 202 (corrected p = 2.0×10^{-5})
- 102 → 202 (corrected p = 2.0×10^{-4}).

Both the uniquely Matthean (201) and the uniquely Lukan (102) subcategories predict the shared vocabulary (202), and the two are not equivalent for the competing hypotheses. The Farrer Hypothesis accommodates 201 → 202 readily: on that view 201 and 202 are alike Matthean in origin, so the diction that links them is unsurprising. But 102 is, on the Farrer view, Lukan redaction — words Luke supplied where Matthew's parallel has something else — and there is no evident reason his editorial vocabulary

should predict Matthew's authored vocabulary in 202 beyond what the corpus baseline already supplies. A common sayings source explains the symmetry directly: 201, 102, and 202 all draw on the one source and all carry its diction, so each predicts the others. The bidirectionality is thus the signature of a shared source rather than of one evangelist copying the other. The most serious counter is shared subject matter: 102 and 202 words may occupy the same pericopes, and topical overlap could in principle lift the prediction without a common source. But the model already subtracts the corpus-wide baseline, so generic topical co-occurrence is absorbed before the test. One might still argue that the sayings genre itself drives the connection — yet even if genre inflated the apparent magnitude of these links, it would not change their *direction*. On the Farrer view 201 → 202 should dominate, since both categories are Matthean in origin, while 102 → 202 — Lukan redaction predicting Matthean vocabulary — should be weak or absent. The data show the reverse of that asymmetry: near-equal weights in both directions (β = .1576 for 201 → 202, β = .1484 for 102 → 202). That symmetry is what a shared source predicts and what the Farrer view does not. The result is therefore strong, if not strictly decisive.

The minor agreements. Two results bear directly on the long-standing problem of the minor agreements — passages in category 212 where Matthew and Luke share vocabulary against Mark. (The larger, "non-minor" minor agreements, such as the Beelzebul controversy discussed below, fall into this category as well.)

- 211 → 212 (corrected $p = 8.0 \times 10^{-4}$) significant at $\varphi = 1.8$, only under FDR at $\varphi = 2.17$
- 210 → 212 (corrected $p = 8.0 \times 10^{-5}$)

Both predictors are categories in which Matthew has the word and Mark does not — that is, distinctively Matthean vocabulary. That this Matthean-signature diction predicts the minor agreements is the feature the standard defenses of the Two-Source Hypothesis do not account for. Those defenses require coincidental identical improvement of Mark's rougher Greek, or overlap between Mark and the sayings source in the relevant passages. If Matthew and Luke had independently polished Mark's rougher Greek and happened to coincide, the shared result would carry no particular author's stamp — it would be the neutral improvement any competent editor might reach for. What the data show instead is Matthew's own distinctive vocabulary surfacing in the agreements (and not Luke's), which coincidental editing has no reason to produce. And Mark–sayings-source overlap fares no better. On that hypothesis the shared material derives from a tradition Mark and the source held in common — yet what predicts the agreements is category 211 and 210 vocabulary, words distinctive to Matthew and absent from Mark.

Overlap with a source Mark also drew on gives no reason for Matthew's own diction to surface in the Matthew–Luke agreements; direct use of Matthew by Luke does. The result is thus not merely in tension with the Two-Source Hypothesis but discriminates between its independent-cause explanations and Lukan use of Matthew, in favor of the latter.

Mark/Luke agreement predicting special Luke. The most unexpected individual result lies outside the hypothesis-testing family (it predicts a Sondergut category) and is reported as exploratory:

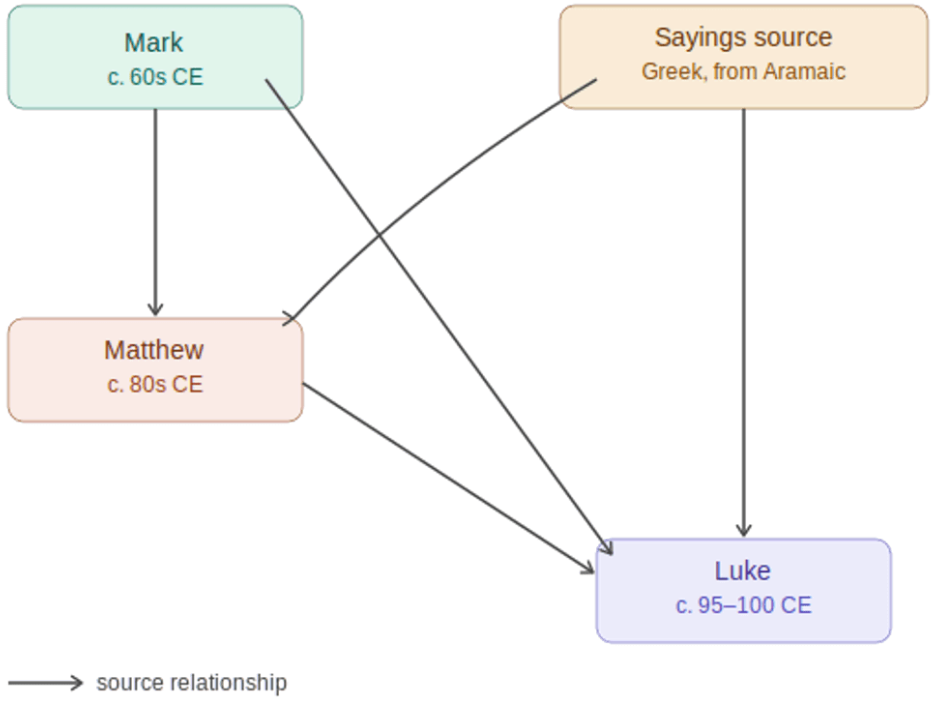
- 122 → 002 (corrected $p = 2.0 \times 10^{-6}$)

Vocabulary shared between Mark and Luke in the triple tradition predicts vocabulary appearing in Luke's Sondergut — material unique to Luke with no Markan or Matthean parallel. The reciprocal relationship does not hold: special-Luke material does not predict the Mark/Luke agreement categories. Examining the details shows that the explanation is compositional: Luke shows a consistent tendency to carry vocabulary encountered in his Markan source forward into his own distinctive material, deploying it again in new contexts. The words θεός and ἄφεσις (forgiveness) are representative examples of this pattern.

6. Discussion

The results establish with a high degree of mathematical confidence that the vocabulary profiles of certain synoptic categories are more closely related to specific individual categories than to the synoptic corpus as a whole. Taken together, they place Markan priority beyond reasonable doubt, provide strong support for a sayings source underlying a large part of the double tradition, and indicate that Luke had access to Matthew, with the balance of the double tradition coming to Luke through Matthew. It is worth adding that the exact degree of significance matters less than the extended discussion above might suggest. The existence of literary dependence among the Synoptics is not in dispute; every hypothesis assumes it. What divides the hypotheses is the *direction* of dependence — and on that question the evidence is even clearer. Because the dispersion correction scales all tests alike, it cannot change which predictor wins: the directional contrast is invariant under any value of φ. And the contrast is stark. 121 → 221, indicating Mark → Matthew, shows up clearly, while nothing like 211 → 221 — which would indicate Matthew → Mark — appears at all. The simplest documentary hypothesis consistent with all of these findings is the Three Source Hypothesis: Mark composed first; Matthew drew on Mark and a sayings

source; Luke drew on Mark, the sayings source, and Matthew. The simplest solution supported by the study is shown in the following diagram:



Attestations that support this hypothesis. Canonical Luke's own prologue (Luke 1:1–4) explicitly acknowledges written predecessors: πολλοὶ ἐπεχείρησαν ἀνατάξασθαι διήγησιν — "many have undertaken to draw up a narrative." Luke positions his own work as a more orderly and complete successor to them. The prologue does not single out any particular source, but it establishes the documentary situation the present hypothesis presupposes: by Luke's time several written accounts were already in circulation, and Luke understood himself to be working from them. Papias supplies a possible glimpse of one such source. He reports that Matthew compiled the λόγια ("sayings" or "oracles") of the Lord in the Hebrew dialect, and that "each interpreted them as he was able" — a notice of a Semitic-language collection transmitted through translation. The term λόγια is disputed, but on one natural reading it refers to a sayings collection rather than to canonical Greek Matthew, and the explicit mention of translation from a Hebrew or Aramaic original fits the loose, translated sayings source the present analysis.

The character of the sayings source. The statistical analysis performed in this study shows that the minor agreements are Matthean in origin. It is likely that Luke borrowed more of the double tradition from Matthew as well, though that is not demonstrated here. The bulk of the minor-agreement material

comes from the largest, most narrative-anchored blocks, such as the Baptist's preaching, the temptation, and the Beelzebul controversy. All three are attested absent from Marcion's Evangelion, consistent with their being later, Matthean additions to the Lukan line rather than part of its early stratum. When this narrative-anchored, Matthew-derived material is set aside, what remains of the double tradition is predominantly freestanding sayings without connected narrative framework. Thus, the sayings source implied by this analysis is not the formal, carefully reconstructed Greek document that Q scholarship has typically envisioned [4]. The evidence is consistent with a considerably more fluid entity: an amorphous collection of sayings material lacking a narrative framework. Its Greek appears to rest at least in part on Aramaic originals: in some passages the Greek preserves traces of an Aramaic wordplay that translation into Greek has partly obscured — a phenomenon examined in the Luke 14 example in the supplement. In overall character the source was likely similar to a sayings collection such as the Gospel of Thomas.

A proto-Luke? The present study does not establish the existence of a proto-Luke, but neither does it preclude one. A possible hint lies in the 122 → 002 result: that Mark/Luke agreement predicts special-Luke material could in principle reflect a late redactor of canonical Luke amplifying vocabulary already present in an earlier Lukan document, carrying the characteristic diction of his source forward into his own additions. This reading is speculative, however. The study shows that the special-Lukan material is lexically derivative of the triple-tradition material in Luke, but it cannot tell us *when* that dependence arose — at the moment of initial composition or in a later redaction many years on. The question of proto-Luke must be argued primarily on other grounds.

The Gospel of the Ebionites, known from patristic citations, appears to have been a Jewish-Christian gospel without birth narrative or infancy material, beginning directly with the adult ministry of Jesus. It had no genealogy, no Lukan prologue, no Magnificat or Benedictus, no nativity sequence. Its profile is precisely that of a gospel in the hypothesized proto-Luke tradition: a document covering the public ministry and passion of Jesus without the elaborate introductory apparatus that canonical Luke supplies. Its existence confirms that gospels of this structural type — beginning with John the Baptist and the adult Jesus rather than with infancy narratives — were circulating in the early period and were not considered incomplete or deficient by the communities that used them.

The best surviving witness to proto-Luke may be Marcion's Evangelion, the gospel used by the Marcionite communities from the mid-second century onward. Marcion, active in Rome in the 140s CE and originally from Sinope in Pontus, used a version of the Lukan gospel that differed from canonical Luke in ways that have generated substantial scholarly debate. The traditional explanation — that

Marcion deliberately excised material he found theologically uncongenial, particularly material tying Jesus to Jewish scripture and covenant history — has been challenged by a series of scholars who argue that the relationship runs in the opposite direction: that the Evangelion represents an earlier form of the gospel which canonical Luke subsequently expanded, and that the large structural absences in Marcion's text are inherited from proto-Luke rather than deliberately removed. [15][16]

I do not want to wade too deeply into this question here, but one argument about the macroscopic shape of the text is worth presenting. If Marcion were responsible for excising large portions of his inherited gospel, it is a considerable coincidence that the result so closely matches the structural profile of the Gospel of the Ebionites — a Jewish-Christian gospel that, on the usual dating, predates his activity and likewise lacked an infancy narrative and a genealogy. This convergence is the more striking because the two communities held irreconcilable views of Christ: the Ebionites regarded Jesus as wholly human, while Marcion's Christology was docetic, granting him only the semblance of flesh. Groups so opposed are unlikely to have been trading gospel texts, which makes independent inheritance of a shared earlier form the more natural explanation for the agreement. The more parsimonious conclusion is that, at least as regards the macroscopic shape of his gospel, Marcion was telling the truth when he said he had inherited it in that form — leaving open, of course, the possibility that he altered details. A reasonable compromise, then, is that both the Evangelion and canonical Luke depend on an earlier proto-Luke whose macroscopic form already resembled Marcion's text.

On this hypothesis — the Three Source Hypothesis with proto-Luke — the expansion from proto-Luke to canonical Luke involved incorporating material Luke drew from Matthew. If Marcion's Evangelion witnesses to that earlier proto-Luke, it should lack precisely those Matthean additions, and on the whole it does. The clearest case is the additional preaching of John the Baptist. That Matthew and Luke insert the same block of Baptist material at the same point in the Markan sequence, in close verbatim agreement, marks it as material Matthew composed and Luke copied directly, rather than sayings-source material the two drew on independently — independent use of a fluid sayings collection would not reproduce both the splice-point and the wording, and the judgment-laden tone of the passage is characteristic of the Matthean stream. On the proto-Luke hypothesis this is a late, Matthean addition to canonical Luke, and the Evangelion's lack of it is exactly what the hypothesis predicts. The Beelzebul controversy (discussed below), the largest and most conspicuous section in which Luke appears to depend on Matthew, is likewise absent. So too are Luke's infancy prologue and resurrection-appearance ending, the impulse for which may itself have come from acquaintance with Matthew. A full treatment of

these agreements could fill a small book; but in the most obvious cases the Evangelion looks like a proto-Luke that has not yet received its late, Matthew-based additions.

The present study cannot adjudicate the Marcion question, but its findings are consistent with the proto-Luke hypothesis and offer it a natural place within the documentary model: an early Lukan gospel, lacking Matthew's contributions, standing behind both the Evangelion and canonical Luke. This offers a natural resolution to one of the more persistent difficulties in Lukan source criticism. If proto-Luke was composed early, drawing on Mark and the sayings source alone, and canonical Luke represents a subsequent expansion of that document informed by Matthew, then the spine of the composition was already fixed before Matthew's influence was felt. This accounts economically for what is otherwise puzzling behavior: rather than following Matthew's Sermon on the Mount as a unified discourse, canonical Luke distributes its contents across widely separated contexts — the Sermon on the Plain in Luke 6, and numerous later insertions scattered through the travel narrative. On the proto-Luke hypothesis this is not arbitrary decomposition but a natural consequence of editorial layering: a redactor working Matthew's material into an already-structured document could not simply transplant the Sermon wholesale, and instead mined it selectively, placing individual units where the existing framework permitted. Use of Matthew as a primary source must attribute this dismantling only to editorial choice, however motivated; the prior framework of proto-Luke renders it close to necessary, requiring no such motive at all. The result is a Lukan gospel dependent on three sources in sequence: Mark and the sayings source at its foundation, by way of proto-Luke, and Matthew as a later, tertiary source shaping the expansion into canonical Luke.

A proto-Mark? Similarly, the study neither establishes nor excludes a proto-Mark, but one feature of the results is worth noting. Markan priority over Matthew emerges considerably more strongly than Markan priority over Luke: the Mark→Matthew relationships clear even the strict Bonferroni threshold, whereas the direct Mark→Luke relationships clear only the false-discovery-rate threshold. The pattern of effect sizes points the same way. The Markan-distinctive predictor of the Mark/Luke agreement carries a larger weight (121 → 122, $\beta = .1777$) than its Lukan-distinctive counterpart (112 → 122, $\beta = .1023$): the dependency appears to run mainly Mark → Luke, though neither weight is individually significant. The smaller but nonzero reverse weight is consistent with — though it does not require — occasional cases in which Luke preserves a form earlier than canonical Mark, as would be expected if both drew on a common earlier document that canonical Mark subsequently revised. Neither weight reaches the significance threshold, so this remains a hint rather than a finding; but the asymmetry between them is a property of the fitted

models, independent of the dispersion correction. The statistical evidence cannot adjudicate the question. The proto-Mark hypothesis is consistent with it, but a full argument is reserved for a forthcoming study [17], and only its relationship to the Three Source Hypothesis is sketched here.

The proposed proto-Mark circulated alongside Paul's ministry in the 40s and 50s CE. Its own author later expanded it into canonical Mark; in its earlier form it also served as a source for proto-Luke, while Matthew used the canonical version. It lacked most of the Great Omission (Mark 6:45–8:26), excepting its closing pericope, the two-stage healing of the blind man at Bethsaida (8:22–26). It also lacked the Beelzebul controversy (3:22–30) and the cursing of the fig tree (11:12–14, 20–25). The appointment of the Twelve (3:13–19) stood in proto-Mark as a choosing of the Twelve with their list of names, the naming of Peter included; the framing apparatus that surrounds it in canonical Mark — the language of formal constitution, the heightening of the naming into a weighted conferral, the second name (Boanerges), and the links to Isaiah 49, to the sending of the Twelve (6:7–13), and to the Beelzebul controversy — belongs to the expansion, which inserted the reframed appointment together with the Beelzebul controversy as a single block. The expansion may have added smaller fragments as well: the greatest commandment (12:28–34), and the anointing at Bethany (14:3–9). However, these lie outside the firm boundaries we ascribe to it.

The rule by which these passages are identified is source-critical, not thematic: canonical Luke attests neither their Markan sequential position nor their Markan wording — the signature of material his source never carried, since on the present model Luke reached the Markan tradition only through proto-Luke and never worked from the expanded gospel of Mark itself. Such expansion material as reached him at all did so only through Matthew's gospel. Applied on these formal grounds alone, the rule yields a set that proves, as a result rather than a premise, to be characteristically pro-Gentile; and the passages so identified show seams internal to Mark. The full reconstruction and that thematic finding are argued in a forthcoming study [17].

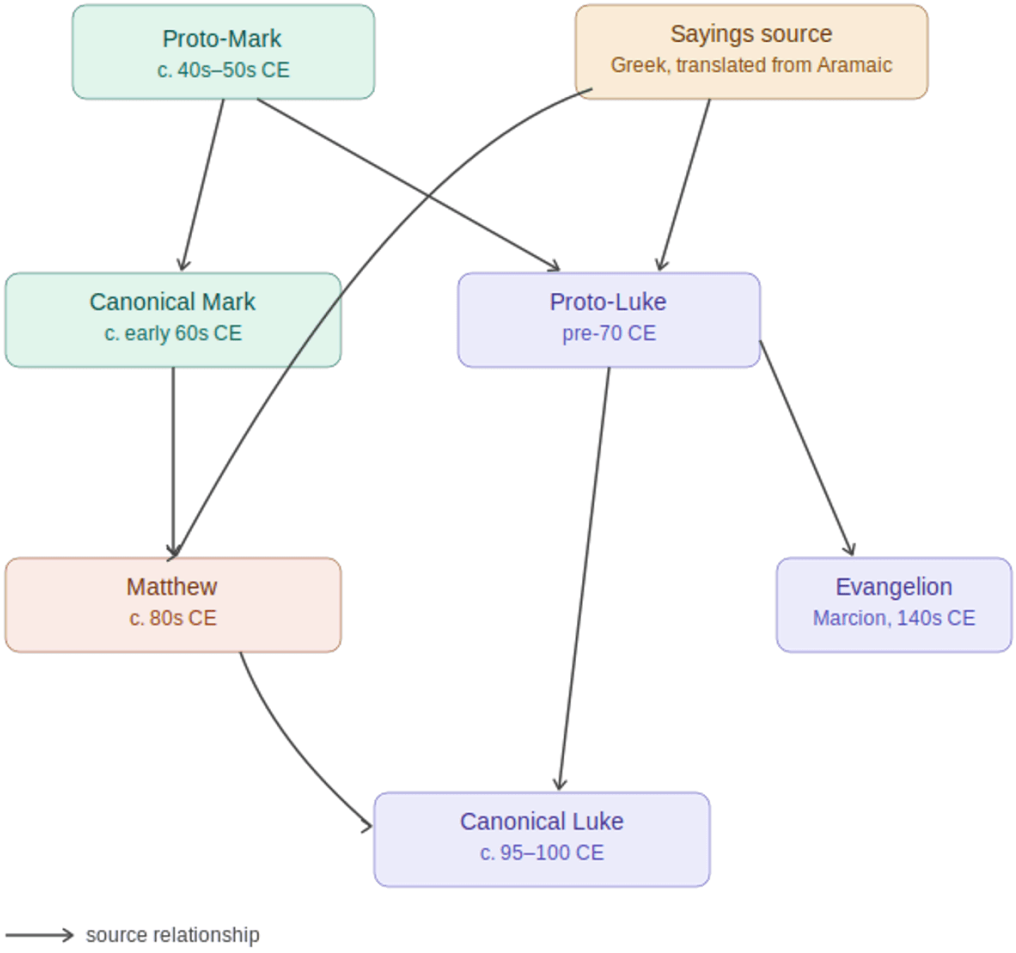
Its existence would complement the Three Source Hypothesis in a way the Beelzebul controversy makes concrete. In this passage Luke shows no knowledge of the controversy's Markan position — he places it in his travel narrative with other non-Markan material — nor does he attest its Markan wording; the two absences one would expect together if the passage were simply not in the Mark he had. He agrees with Mark only where Mark and Matthew themselves agree; at every point where the two diverge, Luke follows Matthew or supplies his own wording. A single substantive exception runs the other way — Luke and Mark read 'Holy Spirit' against Matthew's 'Spirit' — but this is readily explained as Lukan style, the

fuller phrase being a marked Lukan preference attested throughout the Gospel and Acts, rather than as knowledge of Mark′s text.

Advocates of the Two Source Hypothesis must treat this as a Mark–Q overlap; but the reconstructed Q is driven onto Matthew′s own wording — the same pattern seen above in the additional preaching of John the Baptist, where Matthew and Luke agree verbatim against Mark and the reconstructed Q again collapses into the Matthean text. At that point Q has become indistinguishable from Matthew in this passage, and positing it explains nothing that Luke′s direct use of Matthew would not — while still leaving unexplained why Luke followed ′Q′ against Mark′s version of the same episode. If this passage alone were used to solve the synoptic problem, the answer would be straightforward: Mark, then Matthew, then Luke.

The Three Source Hypothesis permits exactly this order of transmission, and the proto-Mark hypothesis explains why Luke′s text looks this way: he never encountered this material in Mark because it was not in the version of Mark he had access to. On any hypothesis that has the passage in Luke′s Mark, we must accept a coincidence — that Luke abandoned both Mark′s order and Mark′s wording in the same passage, when he could have made either change alone. The proto-Mark hypothesis removes the coincidence: there was no Markan version to depart from, because the passage was not in the Mark Luke knew. The Beelzebul controversy is also covered more fully in a forthcoming study [17].

The following diagram summarizes the full documentary model proposed here:



The value of statistical grounding. The broader methodological point is this: the documentary relationships secured here — Markan priority beyond reasonable doubt, a sayings source and Luke's use of Matthew with strong quantitative support — provide a stable foundation from which more speculative proposals about additional documentary layers can be advanced without loss of scholarly footing. Hypotheses about proto-Mark and proto-Luke are not arbitrary if they are constructed as refinements of a model whose primary relationships have been quantitatively secured. The math does not do the work of the literary argument, but it constrains the space within which that argument must operate.

A worked example of all three sources operating within a single pericope — Luke 14:1–6, where Markan narrative, an Aramaic-derived saying from the sayings source, and Matthew's editorial combination of the two can each be traced in Luke's text — is given as an online supplement.

doi.org/10.32388/RIBDZ7.3

7. Conclusions

Three findings carry the argument, and each rests on the same evidential move: which of several candidate predictors predicts a single shared target. That directional asymmetry is a property of the fitted models rather than of the raw p-values, and so survives the dispersion correction intact.

Markan priority is established beyond reasonable doubt. Distinctively Markan vocabulary predicts the material Mark shares with Matthew, while distinctively Matthean vocabulary does not predict that same shared material; the anchor relationship, 121 → 221, remains significant at the order of 10^{-8} after correction and clears even the most conservative dispersion factor tested. Priority over Luke follows from this together with Luke's use of Matthew — a gospel prior to Matthew and used by Luke is necessarily prior to Luke — with the weaker direct Mark/Luke relationships serving as corroboration rather than independent proof. Hypotheses that deny Markan priority, the Griesbach Hypothesis chief among them, face a statistical burden they cannot meet.

The evidence for a sayings source is comparably firm. Both the uniquely Matthean and the uniquely Lukan halves of the double tradition predict the material the two gospels share, with near-equal weight — the symmetry a common source produces, and not what Luke's reworking of Matthew would yield, since his own redactional vocabulary has no reason to track the shared vocabulary so closely. The source so implied is not the stratified Greek document of classical reconstruction but a looser, Thomas-like collection of translated sayings, still bearing in places the Aramaic substrate beneath its Greek.

The minor agreements, long a difficulty for the Two Source Hypothesis, receive a natural account. That distinctively Matthean vocabulary predicts the passages where Matthew and Luke agree against Mark is precisely the trace Luke's use of Matthew would leave; the independent redaction of Mark the Two Source Hypothesis requires cannot produce Matthew's own diction in those agreements.

The simplest documentary hypothesis consistent with all three is the Three Source Hypothesis: Mark first; Matthew drawing on Mark and the sayings source; Luke drawing on all three. This does not foreclose further complexity — the hints toward a proto-Mark and a proto-Luke noted above remain live, and the case for proto-Mark, in particular, is pursued on literary grounds in a forthcoming study. [17] What the statistical analysis supplies is the foundation that makes such proposals disciplined rather than arbitrary: a set of primary relationships secured firmly enough that additional layers can be argued atop relationships not themselves in question. Without that base, further documents introduce

more degrees of freedom than the evidence can constrain; with it, the space the literary argument must work within is fixed.

Appendix A: Definitions of the Synoptic Categories

Each instance of a word in the synoptics is placed into one of a number of synoptic categories, each designated by a three-digit number such as “210.” The first digit gives information about Matthew, the second refers to Mark, and the third to Luke. The digit “0” indicates no parallel. The digit “1” indicates a parallel exists but does not contain the key word. The digit “2” indicates the parallel contains the key word.

222 — Triple tradition, triple agreement

Instances of words occurring in a sentence or clause in Mark, and also in a corresponding parallel in Matthew, and also in Luke. On the 2SH and FH, these are words authored by Mark and copied by Matthew and Luke. On the GH, these are words authored by Matthew and copied by Luke and Mark.

221 — Triple tradition, agreement between Matthew and Mark

Instances of words occurring in Mark and in the corresponding Matthew parallel, but not in the corresponding Luke parallel. On the 2SH and FH, authored by Mark and copied by Matthew but not Luke. On the GH, authored by Matthew, not copied by Luke, then copied from Matthew by Mark.

122 — Triple tradition, agreement between Mark and Luke

Instances of words occurring in Mark and in the corresponding Luke parallel, but not in the corresponding Matthew parallel. On the 2SH and FH, authored by Mark and copied by Luke but not Matthew. On the GH, authored by Luke while editing Matthew, then copied by Mark.

121 — Triple tradition, unique Mark

Instances of words occurring in Mark, but not in corresponding parallels in Matthew or Luke. On the 2SH and FH, authored by Mark but not copied by Matthew or Luke. On the GH, authored by Mark while editing Matthew and Luke.

212 — Triple tradition, minor agreements

Instances of words occurring in Matthew and in a corresponding Luke parallel, but not in the corresponding Mark passage. On the 2SH, authored by Matthew and Luke independently while editing Mark. On the FH, authored by Matthew, possibly copied by Luke or generated independently. On the GH, authored by Matthew and copied by Luke, but not by Mark.

211 — Triple tradition, unique Matthew

Instances of words occurring in Matthew, but not in corresponding parallels in Mark or Luke. On the 2SH and FH, authored by Matthew while editing Mark. On the GH, authored by Matthew but not copied by Luke or Mark.

112 — Triple tradition, unique Luke

Instances of words occurring in Luke, but not in corresponding parallels in Mark or Matthew. On the 2SH, authored by Luke while editing Mark. On the FH, authored by Luke while editing both Mark and Matthew. On the GH, authored by Luke while editing Matthew, not copied by Mark.

022 — Mark/Luke tradition, double agreements

Instances of words occurring in Mark and in a corresponding Luke parallel, where there is no corresponding parallel in Matthew. On the 2SH and FH, authored by Mark and copied by Luke, where Matthew omitted the passage. On the GH, authored by Luke and copied by Mark.

021 — Mark/Luke tradition, unique Mark

Instances of words occurring in Mark but not in a corresponding Luke parallel, where there is no corresponding parallel in Matthew. On the 2SH and FH, authored by Mark and not copied by Luke. On the GH, authored by Mark while editing Luke.

012 — Mark/Luke tradition, unique Luke

Instances of words occurring in Luke but not in a corresponding Mark parallel, where there is no corresponding parallel in Matthew. On the 2SH and FH, authored by Luke while editing Mark. On the GH, authored by Luke and not copied by Mark.

220 — Mark/Matthew tradition, double agreements

Instances of words occurring in Mark and in a corresponding Matthew parallel, where there is no corresponding parallel in Luke. On the 2SH and FH, authored by Mark and copied by Matthew, where Luke omitted the passage. On the GH, authored by Matthew and copied by Mark.

120 — Mark/Matthew tradition, unique Mark

Instances of words occurring in Mark but not in a corresponding Matthew parallel, where there is no corresponding parallel in Luke. On the 2SH and FH, authored by Mark and not copied by Matthew. On the GH, authored by Mark while editing Matthew.

210 — Mark/Matthew tradition, unique Matthew

Instances of words occurring in Matthew but not in a corresponding Mark parallel, where there is no corresponding parallel in Luke. On the 2SH and FH, authored by Matthew while editing Mark. On the GH, authored by Matthew and not copied by Mark.

020 — Sondergut Mark

Instances of words occurring in Mark, where there is no corresponding parallel in Matthew or Luke. On the 2SH and FH, authored by Mark, where both Matthew and Luke omitted the passage. On the GH, authored by Mark, where neither Matthew nor Luke had the passage.

202 — Matthew/Luke double tradition, double agreements

Instances of words occurring in Matthew and in a corresponding Luke parallel, where there is no corresponding parallel in Mark. On the 2SH, authored by Q and copied by both Matthew and Luke. On the FH, authored by Matthew and copied by Luke. On the GH, authored by Matthew and copied by Luke, where Mark omitted the passage.

201 — Matthew/Luke double tradition, unique Matthew

Instances of words occurring in Matthew but not in a corresponding Luke parallel, where there is no corresponding parallel in Mark. On the 2SH, authored by Q and copied by Matthew but not Luke, or authored by Matthew while editing Q. On the FH and GH, authored by Matthew and not copied by Luke.

102 — Matthew/Luke double tradition, unique Luke

Instances of words occurring in Luke but not in a corresponding Matthew parallel, where there is no corresponding parallel in Mark. On the 2SH, authored by Q and copied by Luke but not Matthew. On the FH, authored by Luke while editing Matthew. On the GH, authored by Luke while editing Matthew, where Mark omitted the passage.

200 — Sondergut Matthew

Instances of words occurring in Matthew, where there is no corresponding parallel in Mark or Luke. On all three hypotheses, authored by Matthew in material unique to his gospel.

002 — Sondergut Luke

Instances of words occurring in Luke, where there is no corresponding parallel in Mark or Matthew. On all three hypotheses, authored by Luke in material unique to his gospel.

In summary: the first digit represents Matthew, the second Mark, the third Luke. A “2” indicates the gospel contains an instance of the key word in the relevant clause. A “1” indicates a parallel clause without the key word. A “0” indicates no corresponding parallel clause.

Appendix B: The Hypothesis-Testing Family of Comparisons

This appendix enumerates the full family of 66 directional comparisons to which the multiple-comparison corrections are applied (see the Significance Threshold section). A comparison ″A → B″ (A predicts B) belongs to the family if and only if B is one of the seven cross-gospel agreement categories — 222, 221, 212, 122, 220, 202, 022 — in which the competing hypotheses make differing claims about the authorship and direction of the shared material. For a target with agreement in two gospels, the predictors are the categories carrying the word in exactly one of those two gospels (a ″2″ in one of the target′s agreement positions, with 0 or 1 elsewhere); for the triple-agreement target 222, every other category is a candidate predictor. The rule is content-based and fixed independently of the results.

Predictors of 222 (18): 221, 220, 212, 211, 210, 202, 201, 200, 122, 121, 120, 112, 102, 022, 021, 020, 012, 002.

Predictors of 221 (8): 200, 201, 210, 211, 020, 021, 120, 121.

Predictors of 220 (8): 200, 201, 210, 211, 020, 021, 120, 121.

Predictors of 212 (8): 200, 201, 210, 211, 002, 012, 102, 112.

Predictors of 202 (8): 200, 201, 210, 211, 002, 012, 102, 112.

Predictors of 122 (8): 020, 021, 120, 121, 002, 012, 102, 112.

Predictors of 022 (8): 020, 021, 120, 121, 002, 012, 102, 112.

Categories with agreement in the same two gospels share the same predictor set.

Total: 18 + (6 × 8) = 66 comparisons. The significant members are reported with their p-values in Table 2; the complete set of 66 with all values appears in the accompanying spreadsheet. The relationship 122 → 002 is *not* a member of this family — it predicts the Sondergut-Luke category 002 — and is reported in the Results as an exploratory finding.

Statements and Declarations

Acknowledgments

The author would like to thank all participants in the Synoptic Discussion List for helping this project along in its formative stages, and particularly the list moderators Stephen C. Carlson and Mark S. Goodacre for their encouragement. Thanks also to the late Brian E. Wilson, who assisted with data entry and provided many challenges that helped improve the study.

This study was based on work the author did in 2002, and it has existed on their personal website since then. The author used AI to help clean up the study for publication - specifically Claude (Opus 4.8) was used for copyediting, drafting and condensing expository passages and for consultation on the statistical treatment of overdispersion. The study's conception, its data, its central statistical method, and all scholarly and interpretive judgments are the author's own. The author reviewed and verified all AI-assisted material and takes full responsibility for the content of the study.

References

1. ^*Carlson SC. "Synoptic Problem Website." Synoptic Problem Website. https://www.hypotyposeis.org/synoptic-problem/.*
2. ^*Streeter B (1924). The Four Gospels: A Study of Origins. London: Macmillan.*
3. ^*Tuckett C (1996). Q and the History of Early Christianity. Edinburgh: T&T Clark.*
4. a, b*Kloppenborg JS (1987). The Formation of Q: Trajectories in Ancient Wisdom Collections. Philadelphia: Fortress Press.*

5. ^Goodacre M (2001). The Synoptic Problem: A Way Through the Maze. London: T&T Clark.
6. ^Farmer WR (1976). The Synoptic Problem: A Critical Analysis. 2nd ed. Macon: Mercer University Press.
7. ^Bird M (2012). "The Holtzmann-Gundry Solution to the Synoptic Problem (The Three Source Hypothesis)." Euangelion (blog). Patheos. https://www.patheos.com/blogs/euangelion/2012/05/the-holtzmann-gundry-solution-to-the-synoptic-problem-three-source-hypothesis/.
8. ^Gundry RH (1992). "Matthean Foreign Bodies in Agreements of Luke with Matthew Against Mark: Evidence That Luke Used Matthew." In F. van Segbroeck (Ed.), The Four Gospels. Vol. 2. Leuven: University Press. pp. 1466–95.
9. ^Holtzmann H (1878). "Zur synoptischen Frage" [On the Synoptic Question]. Jahrb Prot Theol [Yearbooks for Protestant Theology]. 4:533–554.
10. ^Price R (2001). "Synoptic Gospel Sources." Synoptic Gospel Sources. https://www.behindthepagesofthenewtestament.uk/syno home.html.
11. a, bHoffmann P, Hieke T, Bauer U (1999-2000). Synoptic Concordance: A Greek Concordance to the First Three Gospels in Synoptic Arrangement. 4 vols. Berlin: de Gruyter.
12. ^Blitzstein JK, Hwang J (2019). Introduction to Probability. 2nd ed. Boca Raton: Chapman and Hall/CRC. http://probabilitybook.net.
13. a, b, c, dGreene WH (2000). Econometric Analysis. 4th ed. Upper Saddle River: Prentice Hall.
14. a, bOakes MP (1998). Statistics for Corpus Linguistics. Edinburgh: University Press.
15. ^BeDuhn J (2013). The First New Testament: Marcion's Scriptural Canon. Salem: Polebridge Press.
16. ^Tyson JB (2006). Marcion and Luke-Acts: A Defining Struggle. Columbia: University of South Carolina Press.
17. a, b, c, dGentile D (2026). "The Great Expansion: Proto-Mark and the Pro-Gentile Enlargement of Canonical Mark." Qeios. doi:10.32388/AWERM6.

Supplementary data: available at https://doi.org/10.32388/RIBDZ7.3

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.