RESEARCH REPORT | 10 SEPTEMBER 2026

Bayesian Structural Inference

Held-out predictive results for the Preedy-encoded Herbal-A sequence

Juan Gabriel Molina
Independent Researcher

Framework by Christine Preedy | Independent testing, reconstruction, and report by Juan Gabriel Molina.

Abstract

This report documents the Bayesian structural-inference experiment associated with the Preedy BSI replication handoff. The primary source contains 16,745 native occurrences in 3,284 blocks across 39 Herbal-A folios. Deterministic encoding produces 16,575 symbols over 28 machine states from a 29-entry key definition. Five folio-grouped folds assess joint held-out predictive probability under a finite, training-derived family of context models, with transition probabilities analytically integrated under symmetric Dirichlet priors. The real primary score is -1.516960 natural-log units per decoded symbol. It exceeds all 500 conditioned M1 twins and all 500 conditioned M2 twins, whose means are -1.743056 and -1.748167. Each empirical upper-tail comparison is 1/501 = 0.001996. The outcome remains unchanged at topology-penalty settings beta = 0.5, 1 and 2, satisfying the package's STRONG_PASS rule. The finding concerns predictive support for this encoded stream relative to the implemented ensembles; it does not establish literal key meanings, unique causal structure, or performance on unexamined manuscript sections. [1-3]

Research question and scope

Does the Herbal-A sequence encoded with the frozen Preedy-29 representation receive stronger held-out Bayesian support than matched conditioned Markov null sequences? The method compares the real stream and synthetic streams using the same scoring pipeline. Bayesian model averaging defines the score; simulation ranks define the reported p-values. These are distinct inferential layers.

The report uses the authors' archived result files linked to the handoff by source, protocol and key hashes. It incorporates no external reviewer grouping tournament, revised null experiment or additional semantic claim.

Keywords: Bayesian structural inference; held-out prediction; context models; Dirichlet evidence; conditioned Markov twins; Herbal-A.

1. Data representation and held-out design

Property	Frozen specification
Source / selection	Included occurrence export; primary_eligible = True
Corpus	39 Herbal-A folios; 3,284 blocks
Native occurrences	16,745
Decoded stream	16,575 symbols; 28 observed machine states
Key definition	29 listed entries; shared r code
Cross-validation	Five disjoint folio groups; all blocks of a folio stay together
Fold seed	20260718

The encoding keeps recognized symbols in its fixed alphabet, groups consecutive e occurrences into e, ee, eee or eeee, and chunks longer runs as eeee plus the remainder. Starred and unusual forms map to SPECIAL. The two claimant r functions are not separately represented as machine states. Consequently, this is not the same 29-state reconstruction used in the structural-null report, which has an explicit no-string-r ledger. [1,2]

The key validator checks required columns, 29 entries, nonblank labels, the expected machine-symbol set and the duplicate-r condition. The decoder itself is implemented in source code. File validation and hashing establish consistency and traceability of the supplied artifact; they do not make the score sensitive to the truth of the English labels.

1.1 Folio folds and weighting

Each folio is held out once. Candidate context sets and their evidence weights are learned from the other four groups. Histories reset at each block boundary. The final score sums fold log-predictive probabilities and divides by the total number of decoded test symbols; it is therefore symbol-weighted, not an unweighted average of five fold scores. [2,3]

Fold	Test folios	Decoded test symbols	Primary score/symbol
1	7	3303	-1.490126
2	8	3286	-1.532021
3	8	3305	-1.531033
4	8	3355	-1.532984
5	8	3326	-1.498581

Table 1. Real-stream held-out folds, numbered 1-5 here; machine outputs use indices 0-4. These folds assess generalization across selected Herbal-A folios, not generalization to other sections or a new transcription.

2. Bayesian scoring model

The package implements a finite context-model adaptation of Bayesian structural inference. The general BSI approach compares candidate unifilar hidden-process topologies using integrated parameter evidence [4]. Here the candidate set is specifically training-derived longest-suffix context models; it is not an exhaustive enumeration of hidden Markov machines or a demonstration that every fitted context state is a minimal causal state. [2]

Component	Implementation
Candidate depths	Root/IID plus maximum context depth D = 1, 2, 3, 4
Support thresholds	Minimum context counts 1, 5, 10, 20, 50
State assignment	Longest retained suffix of current within-block history; root fallback
Duplicate candidates	Identical retained context sets removed
Emission alphabet	K = 28 decoded symbols
Parameter prior	Symmetric Dirichlet, alpha = 1 for every symbol
Topology weight	Prior proportional to exp(-beta × number of retained states)
Primary / sensitivity	beta = 0 / beta = 0.5, 1, 2

2.1 Integrated evidence and prediction

For a fixed topology M, let n(s,a) be the count of symbol a emitted in context state s, and n(s) its total. The integrated likelihood is the product over states of the Dirichlet-multinomial sequence-evidence factors below. No multinomial arrangement coefficient is included because the object scored is the ordered sequence. [2]

E(C | M) = product over s of
[Gamma(K alpha) / Gamma(K alpha + n(s))]
× product over a [Gamma(alpha + n(s,a)) / Gamma(alpha)].

For each fold, the log predictive probability under M is log E(Ctrain + Ctest | M) - log E(Ctrain | M). Training posterior weights are proportional to E(Ctrain | M) exp(-beta |M|). A log-sum-exp calculation averages the joint test probability across candidates using those training weights.

The test portion is scored as a joint posterior predictive sequence. Its context states are fixed by training, but its observed prefixes enter predictive histories, and its counts enter the integrated joint-probability calculation. This is a probabilistic sequence score, not the percentage of glyphs guessed correctly and not a semantic recovery rate.

3. Conditioned null ensembles and decision rule

Native exact-code occurrences are simulated before key encoding. The generator uses smoothed block-start, one-symbol and two-symbol probabilities fitted from the real native stream. Smoothing/backoff strength tau is selected separately for M1 and M2 from {1, 5, 20, 100} by five-fold native-symbol log loss, averaged across folios. Selected values are tau = 5 for M1 and tau = 20 for M2. [1-3]

Within each folio, blocks are visited in a shuffled generation order. At each token position, the next-symbol weight is the fitted start or transition probability multiplied by the number of that symbol still available in the folio inventory. The selected inventory count is decremented. M2 backs off to M1 for unavailable two-symbol contexts. This sequential inventory-depletion procedure is the operational meaning of conditioned twins in this report.

Pchoose(a | history, remaining) is proportional to
PMarkov(a | history; tau) × remaining_count(a).

Preserved by construction	Not asserted to be preserved
Each folio's exact native symbol multiset	Exact bigram or trigram counts
Number of blocks and native block lengths	Exact position-specific symbol frequencies
Folio membership and hard resets	Decoded symbol counts after e-run grouping

Conditioning here should not be read as a proof of exact sampling from an ordinary Markov chain conditioned on its terminal inventory. The finite-inventory sampler is itself the specified null mechanism. Generator parameters are fitted to the real corpus, while Bayesian scoring topologies are rebuilt within the training folds of every real or synthetic stream. Thus the entire simulation design is not an untouched external-corpus validation.

3.1 Simulation count and empirical rank

Each family starts with 100 twins. If the primary upper-tail empirical probability is at most 0.05, it escalates to 500. Both families escalated. The twin seed is 20260907 + twin ID for M1, and that value plus 1,000,000 for M2. Each twin is evaluated at all four beta values.

p = (1 + count of twin scores ≥ real score)/(N + 1).
STRONG_PASS: p ≤ 0.01 for both families, with no qualitative reversal at beta = 0.5, 1 or 2.

These are finite-simulation upper-tail probabilities under the declared adaptive protocol, not posterior probabilities that a key is correct. The report supplies no separate sequential-testing calibration or claim of exact Type-I error control for all data-dependent design choices. The prior settings reuse the same twins and are sensitivity checks, not independent replications.

4. Primary results and prior sensitivity

Family	Real score	Null mean	Null SD	Null 97.5%
M1	-1.516960	-1.743056	0.007060	-1.728565
M2	-1.516960	-1.748167	0.007288	-1.734760

Table 2. Primary beta = 0 score in natural-log units per decoded symbol; higher is better. Each family contains 500 twins, with zero scores at or above real and p = 0.001996. The 97.5th percentile summarizes the null ensemble; it is not a confidence interval for the real score. [3]

The real-minus-null-mean differences are 0.226096 for M1 and 0.231207 for M2. Both are descriptive predictive-score separations. They should not be translated into percentage accuracy gains, probabilities of decipherment, or gains over an alternative key.

beta	Real score/symbol	M1 exceeders / p	M2 exceeders / p
0.0	-1.516960073	0/500 / 0.001996	0/500 / 0.001996
0.5	-1.516934641	0/500 / 0.001996	0/500 / 0.001996
1.0	-1.516935450	0/500 / 0.001996	0/500 / 0.001996
2.0	-1.516960757	0/500 / 0.001996	0/500 / 0.001996

Table 3. State-count-prior sensitivity. The maximum change in the real score across settings is approximately 0.000026 natural-log units per symbol. All four settings satisfy the numerical threshold against both ensembles. [3]

4.1 Descriptive full-corpus posterior

beta	Highest-weight topology	States	Posterior
0.0	CTX_D1_min1	29	0.601820
0.5	CTX_D1_min5	27	0.642664
1.0	CTX_D1_min5	27	0.830186
2.0	CTX_D1_min5	27	0.973063

Table 4. Full-data descriptive fits, separate from held-out scoring. State counts are model contexts, not glyph-alphabet sizes. All listed winners have maximum history depth one. [3]

Allowing histories up to depth four does not establish that deeper memory was selected or necessary. The full-data posterior concentrates on depth-one candidates. The supported conclusion is a predictive-score separation under the specified real/twin comparison, not proof of irreducible long-range syntax.

5. Interpretation, limitations and experiment boundary

5.1 Supported statement

The Herbal-A sequence encoded by the frozen representation receives a higher folio-held-out Bayesian context-model score than every sampled twin from both implemented conditioned ensembles, with stable threshold outcomes across the stated state-count priors. This satisfies the package-defined STRONG_PASS classification. The classification names a protocol outcome, not a universal grade for the broader hypothesis.

5.2 What this experiment does not resolve

The English labels in key_definition.csv are not evaluated against external referents. Altering those descriptions without changing the encoded sequence does not constitute a different semantic test. No matched-size rival-key tournament, plaintext recovery assessment, visual interpretation, or historical attribution is part of this result.

The 28-state representation does not retain the separate no-string-r state of the companion 29-state null experiment. Although both reports use 39 folios and 16,575 encoded units, their source files, encoding details, generators and scores differ. Their p-values therefore cannot be used as a controlled measurement of compression loss or as evidence that one representation outperforms the other.

The scorer uses a restricted, data-derived model family and selected priors. More extensive topology families, different priors, alternative tokenization or more closely matched positional generators could change the comparison. Model averaging within this family does not integrate uncertainty over every possible representation or scientific hypothesis.

Null parameters and the representation are established using the available corpus. Folio-held-out scoring reduces direct reuse of test data in topology fitting, but does not convert the overall study into a fully prospective test. Neither manuscript-wide generalization nor selection of the key before all data inspection is established by the supplied record.

5.3 Relation to the structural-null report

Dimension	Structural-null experiment	This BSI experiment
Representation	29 states; ten-role scoring	28 decoded machine states
Score	Three pooled statistics	Joint held-out log prediction
Generators	M1, M2, finite-history CSSR	Inventory-depletion M1 and M2
Symbol totals	Allowed to fluctuate	Native totals fixed per folio
Target	Aggregate structural mismatch	Predictive-score separation

The two studies provide complementary descriptions of the selected corpus. They share source material and assumptions, so they are not independent semantic confirmations. Repeated deterministic execution tests reproducibility; it does not add independent evidence about meaning.

6. Replication instructions and numerical provenance

Replication packages are available by email request through the contact details accompanying this report on the hosting website. Request the Preedy Bayesian Structural Inference Replication Package. The handoff contains the frozen source, key definition, fold memberships, protocol, implementation, integrity checker and self-test. It intentionally does not bundle prior result files. [1]

From the extracted directory:

```
python -m pip install -r requirements.txt
python verify_integrity.py
python self_test.py
python run_replication.py
```

The BSI_WORKERS environment variable controls process count. The supplied README records verification with Python 3.12.14, NumPy 2.3.5 and pandas 2.2.3. requirements.txt specifies version ranges rather than exact equality pins. Do not infer identical behavior on an untested platform solely from successful installation.

Retain the generated results directory, especially replication_results.json, replication_report.md, twin_results.csv, real_fold_details.json, full_data_model_posterior.json and generator_tau_selection.csv. Runtime metadata can differ across otherwise numerically identical executions. twin_results.csv provides the score-level fingerprint; the result manifest checks the full output bundle.

Handoff ZIP SHA-256
1584bef30b9ff40672f8a3a9d21ce227
5d69f7dbb29d96680e48f565373714c8

Protocol SHA-256
a7e4d73a4caf40c0043c76403eb0fcc8
c3a0e54021e520851b22b2a58bdfffeb

Source occurrences SHA-256
3a73826d94240042923f4b83951cbc4f
770a357a5f1a6bda971a69e67f67380f

Key definition SHA-256
f02d2f27d6bb44c9df1fd037f3236d2f
768fd351d8269c3e8135b25ac4dbdcc4

Archived twin_results.csv SHA-256
5d4b53c4ac387a5b0f99ca26273776bd
36cf2e05394650f731ab0cf5e062ab7c

The source, protocol and key fingerprints above match the authors' archived results. Report preparation verified the handoff manifest, ran its self-test, and checked the 1,000 stored twin rows at all four priors against the reported real scores. No full new ensemble was generated to prepare this report. The self-test is an implementation sanity check, not a proof of statistical validity.

Appendix A. Frozen folds, references and contributions

Fold	Held-out folios
1	f14v, f16v, f1r, f20v, f4v, f5v, f7v
2	f11v, f18v, f19r, f20r, f3v, f7r, f8r, f9r
3	f10v, f16r, f17r, f17v, f18r, f2r, f4r, f5r
4	f10r, f13v, f14r, f1v, f2v, f6r, f6v, f8v
5	f11r, f13r, f15r, f15v, f19v, f21r, f3r, f9v

Table A1. Exact fold membership, taken from the archived real-fold records. Each of the 39 folios appears once. Decoded test-symbol counts and fold scores are in Table 1. [3]

References and source records

[1] Preedy Bayesian Structural Inference Replication Package (2026). Frozen handoff: Preedy_BSI_Replication.zip. README.md; TEST_PROTOCOL.md; METHOD_REFERENCE.md; HANDOFF_INSTRUCTIONS.txt. Original framework: Christine Preedy. Computational testing and packaging: Juan Gabriel Molina.

[2] Same package. run_replication.py; self_test.py; verify_integrity.py; data/key_definition.csv; data/folds.json. These are the operative sources for the model family, encoding, sampling algorithm and checks.

[3] Authors' archived output associated with the matching handoff fingerprints: replication_results.json; twin_results.csv; real_fold_details.json; full_data_model_posterior.json; generator_tau_selection.csv. These records supply the numerical results and fold memberships in this report.

[4] Strelioff, C. C. and Crutchfield, J. P. (2014). Bayesian structural inference for hidden processes. Physical Review E 89, 042119. https://doi.org/10.1103/PhysRevE.89.042119. Preprint: https://arxiv.org/abs/1309.1392. This is the methodological background; the present finite context family, inventory sampler and decision rule are package-specific choices.

Contributions and correspondence

Christine Preedy originated the framework and proposed interpretations. Juan Gabriel Molina contributed independent testing design, computational analysis, reproducibility packaging and report preparation. This report does not claim a journal peer-review status.

For replication materials or questions about the computational procedure, email the report contact listed on the hosting website. Package distribution on request does not substitute for the method, result and limitation disclosures provided here.

External replication and technical review: Edwin Honeycutt reported reproducing the BSI STRONG_PASS result against M1 and M2 under the stated Herbal-A scope. Additional reviewer-proposed tests remain outside this release.