Supplement S1: Appendix F – Automated Scoring Criteria & Algorithm Specification

Relates to: *The Universal Evidence Convergence Framework (UECF)*

1 S1.1 Purpose and Scope

This supplement specifies the fixed, automated binary criteria used to score evidence streams in UECF and the exact algorithm that converts those criteria into convergence scores and tiers. The same automation was used for **all four case studies** in the main text (Exodus, Clovis vs pre-Clovis, Polynesian contact, Prophecy) and for the **30-case backtest** reported in Supplement S3.

A public repository hosts the complete scoring sheets, case study inputs, the 30-case backtest dataset, and reproducibility code:

GitHub repository: `UECF_REPO`
Intended layout: see Section S1.5.
Versioning: reference commits by hash in manuscripts and supplements.

2 S1.2 Automated Binary Criteria (Detailed)

Each metric is scored by fixed yes/no checks. No manual scoring is permitted; automation populates all checks from metadata, text parsing, and citation analysis. Points are summed per metric.

Criterion	Description	Operational pass condition	Pts	Extraction	Not
Empirical Robustness (0–5)					
Peer reviewed	Vetted publication	Venue indexed as peer-reviewed (journal or conference)	+1	Metadata	Whi venu
Independent replication ≥ 2	Reproducibility	At least two distinct groups report confirmatory results	+1	Citation graph	Dist tion
Direct physical evidence	Material or measured	Physical artefact, specimen, instrument reading, or field record	+1	NLP + type tag	Rad stra geno
Confidence $\geq 95\%$	Statistical support	Explicit CI $\geq$95% or equivalently strong test threshold	+1	NLP	Reg valu
Data public	Reanalysis possible	Raw or processed data downloadable	+1	Metadata + link check	Rep supp
Falsifiability (0–5)					
Specific dates/measurements	Concreteness	Contains dated events, locations, or numeric measures	+1	NLP	Dat tors
Testing method stated	Procedure to test	Method/protocol explicitly described	+1	NLP	"Me sect

Criterion	Description	Operational pass condition	Pts	Extraction	Not
Counterfactuals stated	Failure conditions	States conditions under which claim would fail	+1	NLP	Phr "if n
Multiple disprovables	Redundancy	$\geq$ 2 independent parameters could be falsified	+1	NLP + structure	e.g., + a
Independently testable	Third party ready	Third parties can test with accessible tools/data	+1	NLP + metadata	Kit oper
Independence (0–3)					
Distinct authorship/source	Non-overlapping teams	Different first/senior authors or originators	+1	Metadata	Nam bigu
No shared funding	Financial independence	No identical grant IDs or sponsors across key sources	+1	Acknowledgments parse	Gra sors
No top-3 reference overlap	Upstream separation	Top-3 references for two sources differ	+1	Citation graph	Brea upst
Cross-corroboration (0,2,4)					
$\geq$ 2 unrelated categories	Orthogonal agreement	Agreement across categories (e.g., genetic and archaeological)	+2	Type tags + alignment	Cat mai
Independent literature confirmation	Secondary confirmation	Agreement replicated in a separate literature thread	+2	Citation graph	Mus thre

3 S1.3 Algorithm Specification

```
Input:
  Evidence summaries S = {s1,...,sN} and full texts
  Category map, venue whitelist, funder list, repository patterns

For each si in S:
  r = sum([
    is_peer_reviewed(si),
    has_independent_replications(si, k>=2),
    is_direct_physical_evidence(si),
    reports_confidence_ge(si, 0.95),
    has_public_data(si)
  ])                                # r in [0..5]

  f = sum([
    has_specific_measurements(si),
    states_testing_method(si),
    states_counterfactuals(si),
    has_multiple_disprovables(si),
    is_independently_testable(si)
  ])                                # f in [0..5]

  d = sum([
```

```
    is_distinct_authorship(si),
    no_shared_funding(si),
    no_top3_reference_overlap(si, S)
  ])                                # d in [0..3]

  c = 0
  if aligns_with_unrelated_categories(si, S): c += 2
  if confirmed_in_independent_literature(si): c += 2    # c in {0,2,4}

  w = r + f + d + c                    # w in [0..17]
  store (w, d)

Confidence:
  conf = 100 * sum_i(w_i * d_i) / (51 * N)  # 51 = 17 * 3

Tier:
  if conf > 85: VERIFIED
  elif conf >= 60: PLAUSIBLE
  else: SPECULATIVE
```

4 S1.4 Worked Examples (Binary Outputs)

Binary checks (1=pass, 0=fail) for each evidence stream. Cross-corroboration criteria contribute in +2 increments as defined above.

S1.4.1 Exodus Narrative

Evidence Stream	Peer reviewed	Replicated ≥ 2	Physical	Conf. $\geq 95\%$	Data public	Specific dates	Test method	Counterfactuals	Multi disprovables	Indep. testable	Distinct auth.	No shared funding	No top-3 overlap	≥ 2 cats +2	Indep. lit. +2
Biblical Text	1	0	0	0	1	1	1	1	0	1	1	1	1	0	0
Egyptian Records (Absence)	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
Sinai Pottery (1500 BCE)	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Oral Traditions (Twelve Tribes)	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1

S1.4.2 Clovis vs pre-Clovis (First Migration into the Americas)

Evidence Stream	Peer reviewed	Replicated ≥ 2	Physical	Conf. $\geq 95\%$	Data public	Specific dates	Test method	Counterfactuals	Multi disprovables	Indep. testable	Distinct auth.	No shared funding	No top-3 overlap	≥ 2 cats +2	Indep. lit. +2
Radiocarbon dates (> 13 ka)	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Stratigraphic integrity audits	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1
Tool typology comparisons	1	0	1	0	0	1	1	0	0	1	1	1	1	1	0
Origin-motif oral histories	0	0	0	0	0	1	1	0	0	1	1	1	1	1	0

S1.4.3 Polynesian Contact with South America (Pre-Columbian)

Evidence Stream	Peer reviewed	Replicated ≥ 2	Physical	Conf. $\geq 95\%$	Data public	Specific dates	Test method	Counterfactuals	Multi disprovables	Indep. testable	Distinct auth.	No shared funding	No top-3 overlap	≥ 2 cats +2	Indep. lit. +2
Chicken haplotypes (mtDNA)	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Linguistic loanwords	1	0	0	0	0	1	1	0	0	1	1	1	1	1	0
Oral navigation records	0	0	0	0	0	1	1	0	0	1	1	1	1	1	0
Sweet potato dispersal	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1

S1.4.4 Prophetic Migration Exemplar

Evidence Stream	Peer reviewed	Replicated ≥ 2	Physical	Conf. $\geq 95\%$	Data public	Specific dates	Test method	Counterfactuals	Multi disprovables	Indep. testable	Distinct auth.	No shared funding	No top-3 overlap	≥ 2 cats +2	Indep. lit. +2
Evidence Stream	Peer reviewed	Replicated ≥ 2	Physical	Conf. $\geq 95\%$	Data public	Specific dates	Test method	Counterfactuals	Multi disprovables	Indep. testable	Distinct auth.	No shared funding	No top-3 overlap	≥ 2 cats +2	Indep. lit. +2
Textual clause predicates (dated)	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Historical migration registers	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Archaeological strata markers	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Oral traditions (chain of custody)	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1

5 S1.5 Public Repository (GitHub: `UECF_REPO`)

Recommended structure (commit-hash referenced in submissions):

```
UECF_REPO/
  README.md
  LICENSE
  data/
    case_studies/
      exodus_binary_checks.csv
      clovis_binary_checks.csv
      polynesian_binary_checks.csv
      prophecy_binary_checks.csv
    backtest/
      backtest_30cases.csv
  code/
    scoring_pipeline.py
    extract_metadata.py
    parse_citations.py
    compute_confidence.py
    make_roc.py
  outputs/
    s2_extraction_audit_logs/
    s3_threshold_backtest_metrics.csv
    roc_curve.png
```

The S3 figure in the paper can be regenerated by running `code/make_roc.py` to produce `outputs/roc_curve.png`.

6 S1.6 Reproducibility Instructions

1. Clone `UECF_REPO` and checkout the cited commit hash.
2. Place raw PDFs or texts in a local `inputs/` directory if re-extracting; otherwise use the provided binary-check CSVs in `data/case_studies/`.
3. Run `code/scoring_pipeline.py` to compute per-stream metrics and overall confidence:
   - Outputs include per-criterion flags, per-metric sums, (w, d) pairs, and final confidence.
4. To replicate S3, run `code/compute_confidence.py` on `data/backtest/backtest_30cases.csv` and then `code/make_roc.py`.
5. Archive outputs in `outputs/` and record the commit hash in the README.

7 S1.7 Versioning and Auditability

All criteria changes must be versioned. When criteria are updated, rerun Assumption Impact Scoring (AIS) and publish a delta note in `README.md`, including old vs new confidence for each case study and the backtest set.