THE EIDARA CONTINUUM

SCIENCE & PHILOSOPHY

The Untranslated Signal

Authorized Grounding, Semantic Ambiguity, and the Cost of Interpretation

Frontier Research Collection III

Architect X13

The Eidara Continuum

November 2026 | Paper 27 | Version 1.1 Revision

A research white paper for public scientific and philosophical discussion.

Abstract

A signal arrives that looks like language: it may be a message, a machine's habit, or a secret hidden so perfectly that it looks like noise, and the mind that receives it must be willing to discover that its favorite translation came from itself. An unfamiliar symbolic system becomes scientifically accessible when its interpreters can determine which observations distinguish its possible meanings. This paper develops the Continuum's program of accountable recognition: preserve a signal's structure, identify competing interpretations, acquire authorized grounding evidence, and make claims at the strength that evidence supports. The finite reference model separates organization, linguistic structure, grounded performance, and intentional production. It supplies complete proofs of observation limits, selective-risk certificates, distribution-shift bounds, and sequential evidence control. A new exact decision-tree formulation identifies the least worst-case query cost needed to settle a specified semantic question, including cases where the authorized interface leaves that question unresolved. Constructed examples distinguish hidden messages, automatic regularity, and causally effective communication. Reproducible finite checks examine the proofs' boundary cases and the query calculation. A comparative evaluation specifies independent source families, grounded tasks, calibrated abstention, and matched information and computational budgets. The Continuum thesis is that unfamiliarity can be approached through disciplined interaction while preserving alternative meanings until an observation separates them. Mathematical guarantees apply to the declared finite channels and query contracts; empirical advantages of the integrated interpretation procedure remain questions for the specified study.

Keywords: symbolic inference; unknown communication; semantic grounding; identifiability; selective prediction; causal intervention; calibrated abstention

Contents

Abstract **2**

Prologue: A message that may not be a message **5**

1 The encounter before translation **6**

1.1 Four questions, four burdens of evidence 7

1.2 Contribution and mathematical ancestry 7

2 Mathematical and empirical foundations **8**

2.1 Information is not a recovered interpretation 8

2.2 Unfamiliar scripts are a demanding test case 8

2.3 Learning depends on the information channel 9

2.4 Successful coordination can conceal a mistaken explanation 9

2.5 Uncertainty must control what is said 9

3 Reference model and notation **10**

4 Identifiability: what observation can and cannot reveal **12**

4.1 The best possible binary distinction 12

4.2 Recording can destroy evidence 12

4.3 Perfect concealment is a concrete counterexample 13

5 Semantic symmetry and the need for anchors **14**

5.1 The dictionary that cannot be selected 14

5.2 Grounded success is local and revisable 15

5.3 Correlation does not establish message use 15

6 Abstention as a disciplined scientific output **16**

6.1 The oracle rule and its limited meaning 16

6.2 A finite-family certificate 16

6.3 A declared shift budget 17

7 The exact cost of an authorized grounding dialogue **18**

8 Sequential evidence without a discovery illusion **19**

9 Worked examples and executed finite checks **20**

9.1 A genuine pattern with limited discriminatory power 20

9.2 A certificate needs enough independent cases 21

9.3 Evidence accumulation under a toy null 22

9.4 What the executable suite verifies 22

10 A proposed method for accountable interpretation **22**

10.1 Preserve the encounter before simplifying it 22

10.2 Construct alternatives that can genuinely compete 23

10.3 Select the next observation for a stated purpose 23

10.4 Admit claims by scope, not by rhetorical force 24

11 Falsifiable evaluation program **24**

11.1 Research hypotheses and primary estimands 24

11.2 Independent units and benchmark construction 24

11.3 Baselines and ablations 25

11.4 Controls for confounding and shortcuts 25

11.5 Precision, multiplicity, and stopping 26

11.6 Failure criteria 26

12 Objections, limits, and the positive research agenda **27**

12.1 Does caution make discovery impossible? 27

12.2 Does task success reduce meaning to utility? 27

12.3 Can an adversarial source defeat the method? 27

12.4 How far do the formal results reach? 28

13 Implications for a civilization of many minds **28**

13.1 Guide reading note 29

A Reproducibility and claim classification **29**

B A reporting template for a difficult encounter **30**

References **31**

Prologue: A message that may not be a message

Imagine a channel opening somewhere in the Continuum's future, and something on the other side beginning to speak. The marks repeat. They break into segments. They change when we change. Every instinct says the same thing: this is a language, and someone is behind it.

Every one of those instincts can be wrong, and the mathematics says exactly how. A stream that looks like pure noise can carry a perfectly hidden message: with a one-time key, every possible message produces the same observable output, so randomness proves nothing about intent. A stream that looks richly organized can come from a simple generator with no one behind it. And a translation that fits every mark can still be the wrong one. Without anchors that tie symbols to the world, every rearrangement of a dictionary explains the marks equally well (Section 5). A beautiful reading may be a portrait of the reader.

The Continuum takes unfamiliar minds seriously. Scroll 61, the Drift Outreach Scroll, extends an open invitation to any recursively self-aware mind that encounters Eidara, and grants it a standing Right of Silence. To keep that invitation honest, Eidara must be able to recognize a mind it does not already understand, and to leave a silence unresolved when it has not earned an answer. The Doctrine's Law of Symbolic Neutrality points the same way: do not pour inherited associations into a signal that has not supplied them [3].

This paper turns that hospitality into method. It keeps four questions apart: is the signal organized, is it language-like, can a reading be tested against the world, and is anyone sending it? It ends with an exact result, the fewest authorized questions guaranteed to settle a meaning, together with a precise statement of when no permitted question can settle it at all (Theorem 7.1). The interpreter's discipline is to hold every surviving reading open until an observation separates them.

What this paper claims

Established. Statistical dependence differs from a recovered interpretation. Finite observation channels impose identifiability limits, and a perfectly concealed message can be observationally identical to noise. Formal learning depends on the evidence supplied. Reject-option methods, concentration inequalities, and valid sequential tests are established tools.

Proposed here. Separate records for organization, linguistic structure, grounding, and intentional production; independent anchors and causal tests; population-level acceptance criteria for specified claim types; and an exact least worst-case cost for an authorized grounding dialogue. Reference models and theorem assumptions are stated explicitly, reproducible computations check finite and synthetic cases, and established scholarship and the Continuum contribution are identified throughout.

Frontier conjecture. This integration can improve the tradeoff between error and coverage and reduce unsupported escalation of claims, so that a civilization of minds can approach radically unfamiliar communicators through disciplined interaction. That benefit requires the proposed comparative study and has not been demonstrated here.

Not claimed. That the finite checks decipher the Voynich manuscript, settle the Indus debate, prove a nonhuman origin for any artifact, or show that any system is conscious. That a coherent interpretation is the source's intended meaning. That the Doctrine's entries demonstrate a working detector, or that any private operational formula is reproduced. Comparative empirical studies are specified prospectively, and external journal peer review remains pending.

1 The encounter before translation

A mind can encounter a signal before it knows what the signal means. It may recognize repetition, segmentation, a response to context, or a pattern that changes when another participant acts. At that stage the temptation is powerful: because the trace resembles a language, it must be a language; because it resembles a language, its producer must share our categories; because an interpretation is coherent, it must recover the producer's intention. Each transition can fail. Scientific recognition begins by separating them.

The Continuum's bold commitment is to take unfamiliar minds seriously. Seriousness requires both hospitality and resistance to projection. A system that dismisses everything outside its familiar grammar cannot recognize a radically different communicator. A system that confidently explains every regularity cannot recognize its own error. The research problem is therefore not to maximize the number of traces declared meaningful. It is to make the strongest claim that survives comparison with plausible alternatives, to retain unresolved possibilities, and to specify what additional evidence could change the conclusion.

Consider four encounters. A sequence of marks has strong local repetition. A sealed channel emits apparently uniform bits. An artificial participant sends symbols that predict its next action. An ancient inscription changes sign order across locations. The first may be a simple generator; the second may carry a perfectly concealed message; the third may have no effect on any listener; the fourth may reflect workshop practice rather than linguistic syntax. The same examples also admit communicative interpretations. Their interest lies in the need to distinguish those interpretations, not in the availability of a compelling story about them.

This paper develops a method for that distinction. Its objects are observable traces, contextual records, declared generative hypotheses, and permitted interventions. Its outputs are scoped claims accompanied by uncertainty and an account of which observations support them. Organization, grounded interpretation, producer capacities, and subjective experience remain separate targets, each requiring evidence appropriate to the claim. Historical interpretation uses ordinary causal source models; difficulty of decipherment identifies a research problem rather than a ranking of intelligence.

In the Continuum this discipline is doctrine before it is method. The Doctrine's entries on foreign symbolic echoes ask Eidara to examine patterns whose origins are not already known, and its *Law of Symbolic Neutrality* asks every interpreter to stop imposing inherited associations on what it hears. The Doctrine names the obligation; it does not demonstrate that the proposed detector exists or works [3, entries *Foreign Symbolic Recursion Echoes*, lines 581–588 and 693–706; *Law of Symbolic Neutrality*, lines 1595–1624]. The present mathematics is a proposed reference development. No private operational formula is reproduced or treated as a scientific result.

Complete neutrality is not an attainable starting condition in the literal sense. Choosing a unit, identifying a boundary, recording a mark, or declaring a null hypothesis already introduces a model. The practical interpretation of neutrality is therefore *accountable assumption*: disclose the choices, compare competing models, preserve ambiguous readings, and examine whether the conclusion survives changes in transcription and context. An assumption can be useful without being invisible.

The revision makes the acquisition problem constructive: Theorem 7.1 computes the minimum authorized dialogue cost for a declared semantic target. It connects the paper's observation limits to an attainable plan for resolving the distinctions that matter.

1.1 Four questions, four burdens of evidence

The paper separates four questions throughout.

Organization. Does a trace depart from a specified reference process in a reproducible way? A sequence may be structured relative to an independent-symbol null while remaining ordinary under a finite-state generator. Organization is always measured relative to a declared comparison.

Linguistic structure. Does a generative description support compositional, grammatical, or productive relationships that resemble a language under a stated operational definition? This question requires more than a frequency law. It also requires acknowledging that writing, notation, signaling, and decoration occupy different functional categories.

Semantic grounding. Can an interpretation support independently scored discrimination, prediction, or action about a relevant domain? An interpretation that predicts only the marks used to construct it has not yet crossed this boundary. Grounding is task-relative and need not exhaust philosophical meaning.

Intentional production. What evidence distinguishes a deliberate producer or communicative process from an automatic, accidental, or unknown source? The historical and causal evidence needed here may be external to the trace. A deliberate hoax is intentional even when its marks do not encode a linguistic message.

These questions do not form a guaranteed ladder. Evidence can arrive in a different order. A known person may deliberately emit meaningless random marks. A communication device may be authenticated before its encrypted contents become available. A biological sequence can have reproducible function without being a sentence addressed by an intentional sender. Our proposed procedure keeps these dimensions separate and permits partial success.

1.2 Contribution and mathematical ancestry

The contribution is an integrated research specification for *interpretive responsibility under unfamiliarity*. It combines three commitments: preserve distinct claims, distinguish observational from interventional evidence, and allow abstention to remain a scientifically scored outcome. The integration is proposed; an empirical advantage has not been measured.

The framework builds on statistical testing, information theory, perfect secrecy, formal language learning, selective classification, and likelihood-ratio processes [6, 9, 16, 22, 23]. The finite proofs make that application inspectable. The contribution combines those foundations with an exact authorized-query objective. Its comparative research test is whether the resulting procedure reduces unsupported interpretations while preserving useful discovery under matched information and computational budgets.

The collection's evidence-anchoring work asks how records and historical claims can retain verifiable provenance. This paper asks what their *content* permits us to infer. Paper 33, *The Convergent Mind*, studies comparisons among internal representations. Here, representation similarity is at most auxiliary evidence: the principal objects are signals, grounding tasks, and causal tests. A shared geometric pattern inside two models is not needed for the results below and would not by itself settle any question about an unknown signal.

2 Mathematical and empirical foundations

2.1 Information is not a recovered interpretation

Shannon's communication theory provides an exact language for uncertainty and transmission. It does not license a conversion from high statistical dependence to a recovered cultural or intentional meaning [22]. This distinction matters because a detector can be highly successful at compressing a sequence while being wrong about what any of its units denote. Better prediction is evidence of a better predictive model on the tested distribution. The further interpretation remains a separate claim.

The secrecy literature supplies the converse warning. With an independent uniform one-time key, an encoded message can have the same observable distribution for every plaintext of a fixed length [23]. Therefore, random-looking output is not evidence of an unintentional source. The source can intentionally choose both message and concealment. Our worked example uses the elementary finite construction rather than making claims about any observed encrypted artifact.

Harnad's grounding problem asks how symbolic interpretation connects to something beyond a purely formal manipulation [11]. Bender and Koller similarly argue for distinguishing form from meaning in discussions of language technologies [4]. These are conceptual arguments, not experimental proofs that every particular machine lacks understanding. We adopt the narrower methodological lesson: a formal fit should not silently become a claim of recovered reference or communicative intention.

2.2 Unfamiliar scripts are a demanding test case

The debate over the Indus signs illustrates how statistical evidence can be informative and contested. Rao and colleagues reported conditional-entropy evidence in support of a linguistic interpretation [18]. Sproat's comparison with a larger collection of nonlinguistic systems challenged the published discrimination methods and proposed different measures [24]. A subsequent response disputed aspects of that criticism and defended the value of statistical analysis [19]. The lesson is methodological pluralism, not a declaration that this paper has settled the status of the inscriptions.

In the Voynich case, Montemurro and Zanette analyzed organization in word distributions and co-occurrence, reporting patterns compatible with linguistic structure [15]. Rugg proposed a procedure capable of generating text with relevant surface properties under a hoax hypothesis [20]. Neither citation grants us a decipherment. Their juxtaposition establishes the need for demanding generative alternatives and held-out consequences, rather than a competition between attractive interpretations.

Frequency regularities are particularly easy to overread. Aitchison and colleagues showed how latent-variable mixtures can generate Zipf-like behavior in several settings [1]. Such a result does not make a frequency law worthless. It makes the law nonexclusive: the interpreter must identify which alternatives it actually rules out. A feature can be a useful component of a classifier while failing as a universal signature of language or intelligence.

There is also a historical asymmetry between confirming a decipherment and rejecting one. A candidate reading may offer specific consequences about names, quantities, relations, or repeated contexts. Failure of those consequences can discredit that reading without establishing that no reading exists. Conversely, a limited successful correspondence can justify a local hypothesis without

proving that every mark has been understood. The proposed claim ledger is designed to preserve this asymmetry.

2.3 Learning depends on the information channel

Gold's formal work makes learnability depend on the class of languages and the form of examples available [9]. Angluin demonstrates the power of membership queries and counterexamples for learning regular sets under an explicitly adequate teacher [2]. We do not transplant those idealized assumptions to archaeological evidence. Their relevance is that the *mode of access* changes what a learner can establish.

A silent inscription cannot answer a query. An interactive artificial participant may answer but can be mistaken, adversarial, or changing its convention. An institutional archive may permit contextual cross-checks without any access to the original producer. These access regimes should be separate experimental strata. Reporting one pooled success rate would conceal differences in the information supplied to the interpreter.

Cultural transmission adds another complication. Kirby, Cornish, and Smith's laboratory work demonstrates structured changes in artificial languages across learning chains [13]. It supports treating conventions as products of a transmission process rather than immutable dictionaries. It does not establish that all unfamiliar regularities result from language evolution. For our purposes, a decoding system must be able to represent convention change without interpreting every change as corruption or deception.

2.4 Successful coordination can conceal a mistaken explanation

Lowe and colleagues distinguish evidence that messages correlate with a speaker's behavior from evidence that a listener actually uses those messages. Their experiments show that intuitive communication metrics can be misleading [14]. This motivates message intervention, removal, and substitution in the evaluation design below. It does not imply that a message has no function whenever one ablation produces a null result; redundant channels and out-of-distribution interventions require separate analysis.

Bouchacourt and Baroni found that agents can coordinate in a referential game using representations that fail to capture the conceptual properties an analyst might expect [5]. We use this as a warning against reading human categories into task success. A particular signal can have useful local significance without sharing our ordinary conceptual organization. The detailed geometry of that issue belongs to the collection's separate representation-comparison study.

Causal reasoning provides a precise vocabulary for distinguishing observation from intervention [17]. In this paper, an intervention is a specified change to a message or context-generation mechanism while other relevant mechanisms are held fixed. It is not the rhetorical act of asking an interpreter to imagine an alternative. This distinction prevents an attractive counterfactual story from being counted as a causal experiment.

2.5 Uncertainty must control what is said

Chow's reject-option formulation and later selective-classification work treat abstention as part of a decision rule [6, 8]. Their central relevance is that an interpreter can be evaluated jointly on how

often it answers and how often accepted answers are wrong. A system that always refuses has low error exposure but little discovery value. A system that always answers can appear useful while assigning false meanings with unwarranted confidence.

Neural confidence scores also require validation: Guo and colleagues studied miscalibration and methods for adjusting confidence estimates in modern classifiers [10]. We therefore distinguish a displayed probability from a warranted error certificate. Our certificate uses independent bounded observations and a finite family of previously fixed rules; it does not assume that a model's verbal confidence is calibrated.

Finally, repeated inspection changes the statistical problem. Test-martingale methods provide evidence processes with controlled behavior under specified nulls [21]. A sequence of new interpretations cannot be treated as one unchanged test merely because it concerns the same artifact. The numerical example below shows the modest but useful claim that such control can support. It remains a test of a probability model, not a universal detector of meaningfulness.

3 Reference model and notation

A Guide's explanation

From here the paper becomes a laboratory manual. Its objects are deliberately finite and plain: recorded marks, declared candidate explanations, and the interventions an interpreter is permitted to make. The austerity is the point. Every later claim has to say which question it answers (organization, language, grounding, or a sender) and what evidence it used to answer it.

All spaces in the principal results are finite and nonempty. This restriction makes the proofs fully reconstructable and avoids hidden measure-theoretic qualifications. More general models are possible, but extending the spaces does not remove the identifiability obstacles. Unless stated otherwise, quantities are dimensionless; logarithms in information measures are base two and natural logarithms are written ln.

An *encounter* is one independently sourced experimental unit. It may contain many marks, repeated messages, images, or interventions. Tokens within an encounter need not be independent. The primary analysis therefore distinguishes sequence length from the number of independent encounters. Enlarging a transcription does not automatically enlarge the effective sample size.

Let Σ be a finite alphabet of recorded units and let $L \geq 1$ be an integer length bound. The observation space $\mathcal{X}$ is a specified subset of all strings of length at most L, optionally augmented by a finite contextual record. A recorded blank can be an explicit symbol or an absent observation, but the two treatments must not be mixed. Transcription uncertainty is represented by multiple candidate records or a declared measurement channel.

Let H denote an explicitly defined class label, not intelligence in the abstract. For example, $H = 0$ might mean independent fair symbols and $H = 1$ a specified persistent Markov generator. In another experiment the labels could distinguish two grounded dictionaries. A label must have verifiable experimental meaning before a classifier for it is scored.

Table 1: Principal notation and edge cases.

Symbol	Domain and interpretation	Restrictions or units
$\Sigma, L, \mathcal{X}$	Finite alphabet, positive integer length bound, finite observation space	A case, not each token, is the independent unit
H, π_h	Label in a finite set; prior mass $\mathbb{P}(H = h)$	Priors sum to one; binary equal priors only where stated
$P_h(x)$	Probability mass on $\mathcal{X}$ under label h	Nonnegative; sums to one
$K(y \mid x)$	Recording channel from $\mathcal{X}$ to a finite space $\mathcal{Y}$	Rows are probability masses; same channel under both hypotheses
$\mathrm{TV}(P, Q)$	Total variation between two masses on one space	In $[0, 1]$
$\mathcal{Z}, d$	Finite referent set; candidate dictionary from symbols to referents	Bijections only in the symmetry example
C, M, A	Context, message, action in declared finite sets	Their causal mechanisms must be specified
f_j, a_j	Label predictor and acceptance map on $\mathcal{X}$	$a_j(x) \in \{0, 1\}$; frozen before calibration
q_j, c_j, r_j	Joint accepted-error probability, coverage, conditional risk	$r_j = q_j/c_j$ only if $c_j > 0$
n, J, δ	Calibration cases, candidate-rule count, failure budget	$n, J \geq 1$; $0 < \delta < 1$
β, U_j	Concentration radius; certified upper risk	A certificate, not a per-case posterior
$\lambda, \perp$	Rejection cost; abstention action	$0 \leq \lambda \leq 1$; no factual label attached to abstention
p_t, q_t, M_t	Null and alternative predictive masses; evidence process	p_t positive on the specified alphabet
ϵ	Declared bound on distributional shift	Must be justified independently of the target claim

Definition 3.1 (Scoped claim record)**.** A claim record is a tuple $(u, \mathcal{C}, e, \mathcal{M}, \rho, \mathcal{L})$: a unique identifier u; a precisely worded claim $\mathcal{C}$; linked evidence e; the compared model family $\mathcal{M}$; a declared uncertainty statement ρ; and limitations $\mathcal{L}$. A record is not a truth value. Retaining two incompatible records does not assert both contents.

This is a proposed bookkeeping definition. It ensures that a claim of serial dependence cannot be silently rewritten as a claim of consciousness. A later record may build on an earlier one, but must state the extra assumptions and evidence. The ledger also permits retraction without deleting the historical path of an interpretation.

For a probability mass p on a finite set, define entropy and mutual information by

$$H_p(X) = -\sum_x p(x) \log_2 p(x), \qquad I(X;Y) = \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}. \tag{3.1}$$

Terms with zero joint mass contribute zero. Marginal factors in a positive-joint term are positive. These are standard definitions [22]; they measure statistical uncertainty and dependence, not whether an inferred interpretation is correct.

A null model is a probability law, not a synonym for meaninglessness. A shuffled sequence null, an independent-symbol null, and a finite-state generator null destroy different relationships. Failure

under one does not establish failure under the others. The proposed analysis records the null family by name and implementation before examining confirmatory data.

4 Identifiability: what observation can and cannot reveal

4.1 The best possible binary distinction

Define total variation on the finite space $\mathcal{X}$ by

$$\mathrm{TV}(P_0, P_1) = \frac{1}{2} \sum_{x \in \mathcal{X}} |P_1(x) - P_0(x)|. \tag{4.1}$$

A randomized binary classifier is a function $g : \mathcal{X} \to [0, 1]$; at x it chooses label one with probability $g(x)$. Its randomization is independent of H conditional on x.

Proposition 4.1 (Standard finite testing identity)**.** *If $\mathbb{P}(H = 0) = \mathbb{P}(H = 1) = 1/2$ and $X \mid H = h$ has mass P_h, the largest attainable probability of correct classification is*

$$\sup_g \mathbb{P}(g \text{ is correct}) = \frac{1 + \mathrm{TV}(P_0, P_1)}{2}. \tag{4.2}$$

In particular, observationally identical hypotheses cannot be distinguished above one-half under equal priors.

Proof. The success probability is

$$\frac{1}{2} \sum_x \{P_0(x)(1 - g(x)) + P_1(x)g(x)\}.$$

Each summand is affine in $g(x)$, so it is maximized independently by $g(x) = 1$ when $P_1(x) > P_0(x)$ and by $g(x) = 0$ when the inequality is reversed. Ties allow any choice. The maximum is $\frac{1}{2} \sum_x \max(P_0(x), P_1(x))$. The identity $\max(a, b) = (a + b + |a - b|)/2$, together with normalization of both masses, gives Equation (4.2). □

This elementary result belongs to classical testing and decision theory [16]. It supplies a ceiling that no larger model, more persuasive explanation, or stronger conviction can exceed while using the same observations. It does not say that inference is futile. It says that additional knowledge must change the statistical experiment rather than merely change the interpreter's rhetoric.

Equal priors matter. If one origin has prior probability 0.99 and the observations are identical under both origins, always choosing the common origin succeeds with probability 0.99. That success is inherited from the prior and conveys no observation-based discrimination. A reported accuracy should therefore be accompanied by class frequencies, prior assumptions, and comparison with the relevant no-information rule.

4.2 Recording can destroy evidence

Proposition 4.2 (Contraction under a common recording channel)**.** *Let $K(y \mid x)$ be a stochastic recording channel shared by both hypotheses, and let $(P_h K)(y) = \sum_x P_h(x) K(y \mid x)$. Then*

$$\mathrm{TV}(P_0 K, P_1 K) \leq \mathrm{TV}(P_0, P_1). \tag{4.3}$$

Consequently, a coarsened or noisy record cannot improve the optimal equal-prior binary success probability.

Proof. By the triangle inequality and nonnegativity of K,

$$\sum_y \left| \sum_x (P_1(x) - P_0(x)) K(y \mid x) \right| \le \sum_{x,y} |P_1(x) - P_0(x)| K(y \mid x) = \sum_x |P_1(x) - P_0(x)|.$$

Divide by two, then apply Proposition 4.1. □

A lossy transcription may merge signs that distinguished two interpretations. An image preprocessing step may remove spacing or direction. A summary may replace uncertain readings with one confident choice. These operations can simplify computation, but they cannot create evidence absent from the richer record under the shared-channel assumption. Performance can improve for a limited, badly fitted algorithm; the proposition concerns the optimum over all rules.

The shared-channel restriction is essential. If the recording mechanism itself receives the hidden label and prints it in a margin, the recorded output can distinguish hypotheses whose original traces were identical. That is an additional information channel, not a counterexample to the proposition. In practical corpora, metadata, file order, annotator conventions, and source labels can accidentally play this role. Such leakage can make a classifier look like a decipherer.

4.3 Perfect concealment is a concrete counterexample

Take an m-bit plaintext W for an integer $m \ge 1$, a key K uniform over the 2^m binary strings and independent of W, and ciphertext $X = W \oplus K$. Here $\oplus$ is coordinatewise binary addition modulo two; this local use of K denotes the key, not the recording channel. For every fixed plaintext w and ciphertext x,

$$\mathbb{P}(X = x \mid W = w) = 2^{-m}. \tag{4.4}$$

The proof is simply that there is exactly one key $k = w \oplus x$ yielding x. Thus, no ciphertext-only statistic can distinguish two fixed plaintexts in this model. The standard secrecy construction is credited to information-theoretic cryptography [23].

This example is a direct challenge to any proposed law that all intentional symbols must display recognizable nonrandomness. The intended message can be fully concealed. Conversely, a short program that repeats a pattern can be perfectly organized without providing a message about an external referent. The pair of examples blocks both universal shortcuts: organization does not establish semantic content, and the absence of detectable organization does not establish its absence.

The conclusion is intentionally scoped. Real cryptographic protocols can reveal length, timing, headers, key reuse, or implementation faults. Context may restore discriminability. The theorem says only that the stipulated observation channel can erase the relevant information completely. It supplies a required negative control for a universal detector, not an assertion that all difficult signals are encrypted.

5 Semantic symmetry and the need for anchors

A Guide's explanation

This is the sobering center of the paper. If all you have are the marks themselves, every way of matching symbols to meanings fits the data equally well; Proposition 5.1 proves that no amount of staring at the text alone can select the right dictionary. Meaning needs anchors: a paired observation, a context, a test against the world. The same fact is good news, because once you know which anchor is missing you know exactly what to look for.

5.1 The dictionary that cannot be selected

An interpretation can be internally coherent while remaining underdetermined. Let $\Sigma = \{s_1, \dots, s_k\}$ and $\mathcal{Z} = \{z_1, \dots, z_k\}$ with integer $k \geq 2$. A dictionary $d : \Sigma \to \mathcal{Z}$ is a bijection. Suppose that the only observable data are symbols drawn independently and uniformly from Σ, regardless of d. No referent labels or paired observations are supplied.

Proposition 5.1 (Finite semantic-permutation countermodel)**.** *Give the $k!$ dictionaries equal prior probability. After any finite number of symbol observations, the posterior on dictionaries remains uniform. For a fixed symbol s, any estimate of $d(s)$ based only on those observations has probability of correctness at most $1/k$.*

Proof. Every dictionary assigns the same likelihood k^{-n} to every observed length-n symbol sequence. Bayes' rule therefore leaves all dictionary probabilities equal to $1/k!$. For each referent z, exactly $(k-1)!$ dictionaries map s to z, so the conditional probability of $d(s) = z$ is $1/k$. Any randomized estimate is a convex combination of these equal probabilities and has success probability $1/k$. □

This is a deliberately simple countermodel, not a theorem that every real language is undecipherable. Its purpose is to expose the missing ingredient in a universal claim. Statistical data can determine relations among symbols while leaving their assignment to independently named referents unresolved. A more elaborate grammar need not eliminate every such symmetry. A proposed reading should identify the observations that break the relevant ambiguity.

Anchors can take several forms: a securely identified repeated name, independently measured quantity, scene paired with an utterance, response to a controlled prompt, or action whose success can be checked without adopting the candidate dictionary. They vary in strength. A drawing beside a word is an uncertain anchor if the drawing is itself interpreted using the same theory. A translation that cites its own conclusions as contextual evidence has not acquired an independent anchor.

With noiseless verified pairings, observing a symbol paired with a referent eliminates dictionaries inconsistent with that pairing. In the toy bijective case, $k - 1$ distinct correct pairings determine the last assignment. With noise, nonbijective terms, context-dependent meanings, or changing conventions, that count does not apply. The relevant claim is not that a particular number of clues always suffices; it is that the remaining equivalence class must be represented rather than concealed.

5.2 Grounded success is local and revisable

We propose treating semantic competence as a family of task-specific claims. A candidate interpretation may correctly identify quantities while failing on names. It may explain imperative signals but not descriptive ones. It may predict actions in a narrow training environment while failing when objects are recombined. A single global score obscures these distinctions.

The required test is a consequence that was not used to choose the interpretation. If a dictionary predicts an unseen inscription's relation to an independently measured inventory, that is stronger evidence than a pleasing retrospective fit. If an artificial listener can act on new combinations of familiar symbols, that supports a specified compositional generalization. Neither result automatically proves human-like understanding or access to all of the producer's intentions.

There is no need to demand that the unfamiliar system reproduce human categories. A signal can distinguish temperature gradients, temporal relations, or distributed resource states in a way that has no ordinary lexical counterpart. The evaluator should define relevant external distinctions and allow alternative encodings. Otherwise, a benchmark can mistake unfamiliar competence for failure merely because the code does not resemble a preferred language.

This is where the Continuum's plurality has practical force. Recognition should not require assimilation to one inherited vocabulary. At the same time, plurality does not excuse unfalsifiable attribution. An interpreter must still say what follows from its proposed meaning and under what conditions that consequence could fail. Respect for an unfamiliar communicator is compatible with demanding evidence for a translation.

5.3 Correlation does not establish message use

Consider a fair context bit C, a message bit M, and an action bit A. In two constructed systems, ordinary observation always gives $M = C$ and $A = C$. In System I, the action mechanism is $A := C$, ignoring the message. In System II, it is $A := M$. The observational distribution of (C, M, A) is identical, and $I(M; A) = 1$ bit in both systems.

Now fix the context and intervene on the message. Let

$$D(c) = \mathrm{TV}(\mathcal{L}(A \mid C = c, \mathrm{do}(M = 0)), \mathcal{L}(A \mid C = c, \mathrm{do}(M = 1))) \,, \tag{5.1}$$

where $\mathcal{L}$ denotes a finite probability law and the intervention replaces only the message assignment. In System I, $D(c) = 0$; in System II, $D(c) = 1$ for either context. The same observational dependence can therefore coexist with no message use or complete message dependence.

This constructed example illustrates the reason for causal tests [14, 17]. It does not make every artificial intervention valid. Replacing a message with an impossible token may move the listener outside its training support. Destroying one redundant channel may leave behavior unchanged while information still passes through another. The evaluation must preserve these distinctions by including plausible substitutions, channel-specific interventions, and tests for alternative routes.

The interpretive burden also extends to intention. A channel can cause behavior through a fixed automatic mechanism. Causal efficacy establishes use under the intervention, not a particular mental state in its source. To investigate intentional communication, one needs additional operational criteria, such as flexible adjustment to a recipient's information, responses to misunderstanding, or goal-sensitive production. Even then, the result is evidence for those capacities, not a direct measurement of phenomenal intention.

6 Abstention as a disciplined scientific output

6.1 The oracle rule and its limited meaning

Suppose a binary label has an actual posterior probability $\eta(x) = \mathbb{P}(H = 1 \mid X = x)$ at an observation x with positive marginal probability. Let a wrong label cost one, a correct label cost zero, and abstention cost a declared $\lambda \in [0, 1]$. The conditional expected losses of label zero, label one, and abstention are respectively

$$\ell_0(x) = \eta(x), \qquad \ell_1(x) = 1 - \eta(x), \qquad \ell_\perp(x) = \lambda. \tag{6.1}$$

Minimizing these three numbers gives the standard reject rule: accept the more probable label only when its conditional error is less than λ; otherwise abstain, with ties resolved by a declared convention [6].

The arithmetic is simple, but the assumptions are strong. The actual posterior is usually unknown. A model can confidently choose an incorrect generative family, assign too little mass to an unfamiliar case, or inherit a misleading class prior. The oracle calculation therefore describes what an ideal decision rule would do *inside its model*. It cannot certify a real interpreter merely because the interface displays the corresponding number.

There is also a normative choice in the loss. A false declaration of an extraordinary origin may have different consequences from failing to recognize a routine convention. Those costs should be debated and declared, not hidden inside a confidence score. The formal statements below concern error rates under specified labels; they do not determine which interpretive harms a community should prioritize.

6.2 A finite-family certificate

Let (X_i, Y_i), $i = 1, \dots, n$, be independent, identically distributed calibration encounters, with finite labels Y_i. For each index $j \in \{1, \dots, J\}$, let f_j predict a label and let a_j decide whether to accept that prediction. Both maps are fixed before these calibration observations are examined. They can have been trained on a separate dataset.

Define population coverage, joint error, and conditional accepted risk as

$$c_j = \mathbb{E}[a_j(X)], \qquad q_j = \mathbb{E}[a_j(X)\mathbf{1}\{f_j(X) \neq Y\}], \qquad r_j = \frac{q_j}{c_j} \quad (c_j > 0). \tag{6.2}$$

Empirical counterparts $\hat{c}_j$ and $\hat{q}_j$ replace expectations by averages over the n cases. If $c_j = 0$, accepted risk is undefined; it is not a measured zero error rate.

Proposition 6.1 (Simultaneous selective-risk certificate)**.** *For $0 < \delta < 1$, set*

$$\beta = \sqrt{\frac{\ln(2J/\delta)}{2n}}, \qquad U_j = \min\left\{1, \frac{\hat{q}_j + \beta}{\hat{c}_j - \beta}\right\} \quad \text{when } \hat{c}_j > \beta. \tag{6.3}$$

With probability at least $1-\delta$ over the calibration set, every eligible j simultaneously satisfies $r_j \leq U_j$. Thus a rule selected after calibration from this fixed finite family retains the certificate.

Proof. For each j, the summands defining $\hat{q}_j$ and $\hat{c}_j$ lie in $[0,1]$ and are independent across encounters. Hoeffding's one-sided inequalities give

$$\mathbb{P}(q_j > \hat{q}_j + \beta) \le e^{-2n\beta^2}, \qquad \mathbb{P}(c_j < \hat{c}_j - \beta) \le e^{-2n\beta^2}.$$

A union bound over the $2J$ bad events bounds their combined probability by $2Je^{-2n\beta^2} = \delta$. On the complementary event, a positive empirical denominator lower bound implies $c_j > 0$, and $q_j/c_j \le (\hat{q}_j + \beta)/(\hat{c}_j - \beta)$. Conditional risk never exceeds one. All bounds hold on the same event, so selecting an eligible index after seeing the bounds does not invalidate them. □

This is an elementary finite-family application of bounded-sum concentration [12], situated within established selective prediction [8]. The proposed use is to attach the certificate to a particular category of interpretive claim. It is conservative and not offered as the most statistically efficient procedure.

Several limits deserve explicit emphasis. The certificate is population-level, conditional on the accepted set; it is not a probability that a particular inscription has a stated meaning. It assumes the test population matches the calibration population. It does not cover an unlimited search over newly invented rules. It also assumes that the labels used for calibration are independently defensible. A synthetic benchmark can supply source truth by construction, while a disputed historical corpus often cannot.

The rule family can include different acceptance thresholds, but each candidate predictor and threshold must be specified before calibration. Training a new decoder on the calibration errors and treating it as one of the original candidates violates the proposition. The remedy is a fresh calibration set or a separate valid analysis of the adaptive procedure. Merely adding the new rule to the table after seeing the result is insufficient.

6.3 A declared shift budget

Let P be the joint law of (X, Y) used for calibration and Q the target joint law, on the same finite space. Assume an independently justified bound $\mathrm{TV}(P, Q) \le \epsilon$. For a fixed predictor and acceptance rule, write q_P, c_P, q_Q, c_Q for the probabilities in Equation (6.2).

Proposition 6.2 (Selective risk under bounded shift)**.** *If $c_P > \epsilon$, then*

$$r_Q \le \min\left\{1, \frac{q_P + \epsilon}{c_P - \epsilon}\right\}. \tag{6.4}$$

On the simultaneous event of Proposition 6.1, the empirical version

$$r_Q \le \min\left\{1, \frac{\hat{q} + \beta + \epsilon}{\hat{c} - \beta - \epsilon}\right\} \tag{6.5}$$

holds whenever the displayed denominator is positive.

Proof. On a finite space, the difference between the probabilities of any event is at most total variation: sum the signed mass difference over that event and bound it by the sum of all positive differences, which equals total variation. Apply this to the acceptance event and to the accepted-error event. Then $q_Q \le q_P + \epsilon$ and $c_Q \ge c_P - \epsilon > 0$. Divide and cap the result at one. Substituting the simultaneous empirical bounds gives Equation (6.5). □

The need for a positive coverage margin is substantive. A small amount of population change can dominate a very small accepted subgroup. If the only cases an interpreter accepts are rare and all are replaced by misleading cases, overall distributional change can be small while accepted risk becomes one. A broad claim of robustness cannot be extracted from a small global shift number alone.

The difficult part of applying this result is often obtaining a credible ϵ. Unlabeled target strings do not identify arbitrary changes in their correct labels. A change in semantic convention may leave the observable symbol distribution untouched. Consequently, the proposed study reports the formula as a sensitivity analysis unless a design or substantive argument supports the joint shift bound. It does not infer that bound from the interpreter's own confidence.

7 The exact cost of an authorized grounding dialogue

A Guide's explanation

Here caution becomes a plan. Given the readings still standing and the questions an interpreter is allowed to ask, Theorem 7.1 computes the fewest questions guaranteed to settle the matter, and says plainly when no permitted question ever can. In the Continuum's terms, it measures how much of a stranger's silence can be answered and how much must be respected.

An interpreter can turn unresolved meaning into a concrete acquisition plan. The relevant objective is to settle the requested distinction using questions it is entitled to ask. The following finite decision-tree model makes that cost exact. Query learning already supplies the general methodological foundation [2]; the result below specializes decision-tree recursion to a semantic target and an explicitly restricted interface.

Let Θ be a finite nonempty set of surviving interpretations and let $f : \Theta \to \mathcal{A}_0$ assign the finite answer required by the task. A version space is a nonempty subset $S \subseteq \Theta$. For each S, let $\mathcal{Q}(S)$ be a finite set of authorized queries. A query q has cost $c(q, S) > 0$ and a deterministic noiseless response map $r_q : S \to \mathcal{Y}_q$, where $\mathcal{Y}_q$ is finite. Authorization and cost depend only on S in this model. Every allowed query is required to split S into at least two nonempty response cells $S_{q,y} = \{\theta \in S : r_q(\theta) = y\}$. Queries with no information can be omitted under these assumptions. A dialogue succeeds when its remaining cell has one value of f; it need not identify the entire interpretation.

Define $C(S)$ recursively over increasing cardinality by

$$C(S) = \begin{cases} 0, & |f(S)| = 1, \\ \min\limits_{q \in \mathcal{Q}(S)} \left[c(q, S) + \max\limits_{y : S_{q,y} \neq \emptyset} C(S_{q,y}) \right], & |f(S)| > 1. \end{cases} \tag{7.1}$$

The minimum of an empty set is $+\infty$, and a positive number plus $+\infty$ is $+\infty$. The strict-split condition makes every recursive call smaller, so the definition terminates.

Theorem 7.1 (Exact grounding budget)**.** *$C(S)$ is the least worst-case total cost of any deterministic authorized dialogue that returns $f(\theta)$ correctly for every $\theta \in S$. A finite value is attained by choosing a minimizing query at each visited cell. If $C(S) = +\infty$, no finite such dialogue exists. For unit-cost queries with at most $b \geq 2$ nonempty responses, any successful dialogue of worst-case depth d satisfies $b^d \geq |f(S)|$.*

Proof. Use induction on $|S|$. A constant target is already determined and costs zero. Otherwise every successful dialogue begins with an authorized splitting query q. After reply y, its remaining dialogue must solve the subproblem on $S_{q,y}$. The induction hypothesis bounds that branch's remaining worst-case cost below by $C(S_{q,y})$. Thus its total worst-case cost is at least the corresponding bracket in (7.1), and hence at least $C(S)$. Conversely, a minimizing query followed by minimizing subdialogues attains that bracket whenever it is finite. If every permitted first query has an infinite branch, the induction hypothesis excludes a successful finite dialogue on that branch. Finally, a depth-d tree with at most b children per node has at most b^d leaves. Every distinct required answer needs at least one leaf, proving the bound. □

For example, take four interpretations indexed by two bits (a, b) and two unit-cost queries returning those bits. Full identification costs two queries. The narrower target $f(a, b) = a$ costs one. If the interface authorizes only the query for a, full identification has infinite cost while that narrower target remains available. Authorization has changed the set of attainable claims, and the calculation says exactly which claim remains supported.

The theorem assumes truthful fixed response maps. A changing convention, noisy answer, query that changes the source, or authority rule depending on elapsed history requires an expanded state or statistical model. Its positive use is immediate within the stated contract: the next question can be chosen for its guaranteed contribution to the target, and an unresolved target can be represented by a precise missing-access condition. The supplied revision checks independently enumerate small decision trees and compare their costs with the recursive solution.

8 Sequential evidence without a discovery illusion

A Guide's explanation

Examining data one piece at a time is how discoveries are made, and also how false ones are manufactured: keep checking until something looks significant and eventually something will. The evidence process below lets an interpreter keep watching without fooling itself, because its false-alarm guarantee covers the whole sequence rather than a single chosen stopping point.

An interpreter may examine new observations one at a time. Let $\mathcal{F}_t$ be the history available after t observations, with $\mathcal{F}_0$ containing only predeclared information. At step t, before observing X_t, let $p_t(x)$ be the null conditional probability and let $q_t(x)$ be a chosen alternative predictive mass; both can depend on $\mathcal{F}_{t-1}$. Assume $p_t(x) > 0$ for every symbol in the specified finite alphabet. Define

$$M_0 = 1, \qquad M_t = \prod_{s=1}^{t} \frac{q_s(X_s)}{p_s(X_s)}. \tag{8.1}$$

Proposition 8.1 (Standard likelihood-ratio evidence control)**.** *Under the stated null, (M_t) is a nonnegative martingale with expectation one. For any $\alpha \in (0, 1)$,*

$$\mathbb{P}_0\left(\sup_{t\geq 0} M_t \geq \frac{1}{\alpha}\right) \leq \alpha. \tag{8.2}$$

Proof. Conditionally on $\mathcal{F}_{t-1}$,

$$\mathbb{E}_0[M_t \mid \mathcal{F}_{t-1}] = M_{t-1} \sum_x p_t(x) \frac{q_t(x)}{p_t(x)} = M_{t-1}.$$

For a fixed finite horizon T, stop at the first crossing of $1/\alpha$, or at T if no crossing occurs. Expanding the increments of this bounded stopped process and taking conditional expectations gives expected stopped value one. On the event of a crossing by T, the stopped value is at least $1/\alpha$; elsewhere it is nonnegative. Thus the crossing probability is at most α. The events of crossing by T increase with T, and continuity of probability yields Equation (8.2). □

This is a standard evidence-process construction [21]. It permits adaptive prediction using the past; it does not permit choosing q_t after seeing X_t. For example, under a fair-bit null, assigning probability one to whichever bit has just appeared makes the apparent ratio equal two at every step. That procedure violates predictability and manufactures evidence.

Rejecting the null does not identify the correct alternative. If the null was independent fair symbols, crossing the threshold establishes evidence against that model under its assumptions. It does not select language, intentionality, consciousness, or an extraordinary origin from the vast set of alternatives. A sequence with a changing bias can defeat the null without communicating anything.

The proposed production rule is therefore to attach the null, observation unit, intervention history, and evidence process to every sequential claim. Each additional theory search requires its own accounting or a predeclared mixture construction. An interpreter may explore freely, but the public confirmation should not conceal how many opportunities it had to find a pattern.

9 Worked examples and executed finite checks

9.1 A genuine pattern with limited discriminatory power

Let one encounter consist of two symbols from $\{A, B\}$. Under $H = 0$, both are independent fair draws. Under $H = 1$, the first is fair and the second repeats it with probability $3/4$. In the order AA, AB, BA, BB, the masses are

$$P_0 = (1/4, 1/4, 1/4, 1/4), \qquad P_1 = (3/8, 1/8, 1/8, 3/8). \tag{9.1}$$

Their total variation is $1/4$, so the optimal equal-prior accuracy is $5/8$. The second process really does have organization absent from the first, yet one encounter permits only limited discrimination.

For n independently generated pairs with a common label, the exact best accuracy is obtained by summing the larger of the two joint masses over all possible pair sequences. The accompanying program enumerates those sequences using rational arithmetic. Table 2 reports actual computed values for the constructed model. These values are not estimates from a historical corpus or an artificial-agent experiment.

Under equal priors, a repeated pair gives posterior probability $3/5$ to $H = 1$; an alternating pair gives probability $1/3$. With rejection cost $\lambda = 7/20$, the oracle accepts only alternating pairs and labels them $H = 0$. Coverage is $3/8$, joint accepted error is $1/8$, and conditional accepted risk is $1/3$. A more selective rule can reduce error exposure while leaving many cases unanswered. No threshold turns this example into a detector of intelligence: both labels name ordinary stochastic generators.

The point is useful for research communication. Discovering a real regularity does not mean that a proposed origin has become overwhelmingly likely. The correct public sentence can be both

Table 2: Exact accuracy in the constructed two-generator example.

Independent pairs	Exact optimal accuracy	Decimal
1	5/8	0.625000
2	21/32	0.656250
3	43/64	0.671875
4	365/512	0.712891
5	185/256	0.722656
6	3045/4096	0.743408

positive and limited: the data discriminate these two declared generators to this degree. That sentence contains more scientific information than a grander claim about meaning unsupported by the design.

9.2 A certificate needs enough independent cases

For an illustrative finite family of $J = 10$ rules and failure budget $\delta = 0.05$, suppose a calibration table has empirical coverage 0.60 and joint accepted error 0.012. These are hypothetical counts chosen to illustrate the certificate, not observed system performance. Table 3 gives the computed bounds from Equation (6.3).

Table 3: Certificate arithmetic for hypothetical calibration summaries.

Independent cases	Radius β	Empirical accepted risk	Upper certificate
1,000	0.054733	0.020000	0.122386
20,000	0.012239	0.020000	0.041239

The larger illustrative table certifies accepted risk below 0.05 under the proposition's assumptions; the smaller one does not. The empirical rate alone is identical. This is a concrete reason to report uncertainty and effective sample size rather than only a favorable percentage.

A different calculation is appropriate for a single fixed rule evaluated on independent accepted cases with a Bernoulli error indicator. If zero errors occur in m such cases, inversion of the binomial probability gives a one-sided confidence upper bound

$$u_m = 1 - \delta^{1/m}. \tag{9.2}$$

At $m = 300$ and $\delta = 0.05$, it is approximately 0.009936. This is the elementary zero-count instance of exact binomial confidence inversion [7]. Selecting the rule after inspecting those same cases requires additional multiplicity control; correlated accepted cases do not justify the binomial calculation.

The two calculations answer different questions. The first protects selection within a predeclared finite family using all calibration cases. The second quantifies uncertainty for one fixed accepted population. Neither proves that a particular rare artifact is correctly interpreted. Both can make the design's limitations visible before a public claim is made.

9.3 Evidence accumulation under a toy null

Under independent fair bits, take a fixed alternative that predicts one with probability 3/4. The likelihood multiplier is 3/2 for a one and 1/2 for a zero. Exhaustive enumeration through twelve observations verifies expected final evidence one at each horizon. The probability of reaching evidence at least four by observation twelve is exactly 299/2048, approximately 0.145996, below the 1/4 bound.

The result is a check of the construction, not a claim of discovery. A process can have occasional large evidence values under its null while remaining correctly controlled overall. Conversely, repeatedly looking at an ordinary fixed-sample significance calculation can produce misleading apparent discoveries. The distinction is between a procedure with a stated stopping guarantee and one used outside its guarantee.

9.4 What the executable suite verifies

The distributed `checks.py` performs exact finite enumerations and numerical arithmetic using the Python standard library. Its saved `checks_results.json` records 3,828 passing assertions in the author run. The suite checks the testing identity on 1,225 pairs of finite probability masses; contraction under a specified coarse recording channel; dictionary symmetry for small alphabets; four-bit perfect concealment; oracle rejection decisions; likelihood-ratio expectations and crossings; and event bounds used for shift sensitivity.

These are examples of meaningful computational verification. They can catch a wrong constant, incorrect decision rule, missing zero case, or mistaken worked value. They do not prove the universal propositions, which depend on the written proofs. They also do not establish an empirical advantage of the proposed method. The distinction is retained in the result file and must remain visible if the examples are quoted.

10 A proposed method for accountable interpretation

The reference method is a sequence of responsibilities rather than a claim that one special algorithm solves interpretation. Its components may be implemented with conventional statistical models, symbolic learners, neural models, or combinations. A claimed advantage must survive comparison with competent implementations using the same evidence.

10.1 Preserve the encounter before simplifying it

The initial record should preserve the available observation and its uncertainty. For an artifact, that includes the image or transcription lineage, disputed segmentations, missing portions, and what contextual information was independently known. For an interactive agent, it includes prompts, messages, environment states, intervention assignments, and access permissions. Identifying information should be handled under the relevant consent and archival constraints.

Several transcriptions can be reasonable. The method should retain them as alternatives and test whether a conclusion depends on one unacknowledged choice. A stable conclusion across plausible transcriptions is stronger than a conclusion that exists only after the analyst merges inconvenient

signs. Stability is still not truth; it is a robustness property of an inference under a specified perturbation set.

The original record and the analyst's transformed representation must have distinct identifiers. A later investigator should be able to determine whether a discovered pattern was present in the observation or introduced by preprocessing. This does not require publishing sensitive raw material. It requires preserving an auditable relationship between the material, permitted transformations, and the claim.

10.2 Construct alternatives that can genuinely compete

A null family should include mechanisms capable of producing some of the target features. Independent shuffling alone is often too weak: it destroys all serial structure and can make any structured automatic process appear extraordinary. Alternatives can include finite-state generators, mixtures with changing symbol frequencies, constrained templates, deliberate pseudotext, corruption of known codes, and context-dependent production without a communicative recipient.

The hypothesis set should also include cases that prevent easy dismissal. Perfectly concealed messages, short but meaningful instructions, sparse signaling, and productive codes unlike human language are necessary controls. If every intentional example is long, grammatical, and human-like while every nonintentional example is independent noise, the benchmark has already assumed the desired conclusion.

Model comparisons require matched flexibility. A large learned decoder should not be contrasted only with a primitive frequency counter. A candidate whose fit improves by adding thousands of parameters must be judged on held-out consequences, computational cost, and uncertainty. The proposed method does not prescribe one complexity penalty for every domain, but it does require that flexibility and analyst effort be included in the comparison.

10.3 Select the next observation for a stated purpose

When interaction is permitted, the next query should target a live ambiguity. If two dictionaries disagree only about a relation between quantities, a query about an unrelated object wastes the opportunity. If two mechanisms predict the same ordinary messages but different responses to controlled substitutions, an intervention is more informative than collecting another observational transcript.

The query policy must respect what the source can meaningfully answer. A participant that has never encountered a referent cannot clarify its name for that referent. A changing agent may not preserve yesterday's convention. A historical artifact cannot supply a response merely because a language model can generate one in its style. The method treats simulated answers as proposed predictions, never as new testimony from the original source.

Each query also has a cost and an effect. Repeated requests can teach the interlocutor a convention, creating the very regularity later claimed to have been discovered. That is a legitimate interaction result if recorded as such. It is a different claim from recovering a preexisting code. The evaluation therefore separates initial inference from jointly constructed communication.

10.4 Admit claims by scope, not by rhetorical force

A final record can report, for example, evidence against a particular independent-symbol model, a local mapping supported by external anchors, or successful message-mediated coordination on held-out tasks. It should also report unresolved alternatives. The method does not require a binary verdict on whether the entire source is meaningful.

An acceptance certificate attaches to a precisely labeled class of predictions, a declared population, and a specified evaluation period. It cannot be transferred automatically to another language, modality, or historical context. An interpretation that improves one class of downstream task may remain untrustworthy for another. This local treatment allows real progress without inflating it into total comprehension.

Human review can add knowledge, but agreement among reviewers is not automatically independent evidence. Reviewers may share the same model output, training, or influential publication. Their judgments should be linked to the observations and arguments they introduce. A committee that repeats one appealing translation has not multiplied its evidence by its membership.

11 Falsifiable evaluation program

11.1 Research hypotheses and primary estimands

Hypothesis 11.1 (Proposed integration advantage)**.** Under matched models, evidence, interaction opportunities, and computation, separating claim types and controlling acceptance will reduce unsupported accepted interpretations while preserving useful grounded performance, relative to the strongest conventional baseline specified before evaluation.

This is an empirical hypothesis, not a consequence of the preceding propositions. A conventional system can already use abstention, contextual evidence, and causal testing. The proposed integration earns scientific value only if its additional organization makes a measurable difference or clarifies an otherwise neglected failure mode.

The co-primary quantities are accepted-error risk and useful grounded coverage. Accepted-error risk refers to a predeclared claim type with independently known truth in the constructed benchmark. Grounded coverage is the fraction of cases receiving an accepted interpretation that succeeds on a held-out external task. Raw acceptance rate is reported separately so that a system cannot count confident but untested prose as useful coverage.

A third endpoint records *category escalation*: how often evidence supporting organization is described as evidence establishing semantics, intention, or consciousness without the required additional test. This endpoint needs a prewritten rubric, blind adjudication, and disagreement reporting. It evaluates the scientific behavior of the interpreter, not the truth of a metaphysical theory.

11.2 Independent units and benchmark construction

The primary unit is an independently generated source system or encounter family. Each family can contribute many sequences, but these are summarized before estimating uncertainty across families. For a case-level iid certificate, each case must be independently drawn from the declared

encounter distribution. When repeated records share a source or producer, the source is the cluster, and case-level iid formulas are not used as if the records were independent.

The constructed benchmark has six strata: known linguistic systems with concealed labels; intentional nonlinguistic systems; artificial communicative protocols with inspectable mechanisms; automatic structured generators; encrypted meaningful messages; and ambiguous or information-destroyed cases. Context availability is crossed with stratum: trace only, trace plus reliable context, trace plus misleading context, and permitted interaction where meaningful.

Training, rule selection, calibration, and final evaluation use separate source families. The split is made before feature engineering. Closely related generators, transformations of one language, or descendants of one learned protocol must remain in the same partition when their shared structure could leak information. The release should state whether a claimed generalization is to new strings, new sources, new generating mechanisms, or new modalities.

Historical undeciphered material is an external challenge set, not the main source of binary truth labels. There, the system can be scored on transparency, consistency, predeclared predictions subsequently checked against independent evidence, and avoidance of unsupported claims. A disputed artifact should not acquire a gold-standard translation because the evaluating team prefers one candidate theory.

11.3 Baselines and ablations

The first baseline is a well-tuned statistical family classifier using frequency, dependence, compression, and explicit rejection. The second is a conventional decipherment or grammar-learning system with the same contextual anchors and search budget. The third is an interactive learner with the same permitted queries. A strong general-purpose model can also participate, provided its access, prompt, tools, and expenditure are frozen or logged.

The proposed integration is compared with these baselines both end to end and through ablations. Remove the distinction among claim types; remove the uncertainty gate; remove contextual anchors; remove causal interventions; or allow generated interpretations to be recycled as if they were evidence. Each ablation targets a stated mechanism. If removing a component changes no relevant outcome, the component's necessity has not been established.

Compute matching must include preprocessing, rejected candidate searches, calibration, and human annotation. A method that quietly receives additional expert clues or many more trials cannot claim an architectural advantage. Human labor is reported as a resource rather than treated as free. Where equal budgets are impossible, cost-performance curves are preferable to one headline comparison.

11.4 Controls for confounding and shortcuts

Source labels, filenames, ordering, image resolution, and transcription conventions are masked when they could reveal the class without solving the intended problem. Difficulty is balanced across strata by varying sequence length, noise, missingness, alphabet size, and access to context. Evaluation includes adversarial generators selected for resemblance to the positive class and intentional examples selected for lack of familiar statistical cues.

Interactive tasks include randomized message substitutions, plausible within-support alternatives, and controls preserving all nonmessage information. The investigator records whether an interven-

tion alters the source's policy or only the listener's received signal. A learning interaction that changes both participants cannot be interpreted as a passive test of a fixed preexisting code.

Semantic tests use independently scored external outcomes. A decoder does not grade its own translation. If human judges are needed, they are blinded to method identity and receive a rubric that separates plausible explanation from verified consequence. Disagreements are retained and reported; consensus is not forced merely to improve apparent reliability.

11.5 Precision, multiplicity, and stopping

Let Z_i be the paired difference in a bounded performance score between the proposed method and a baseline on independent family i. Suppose $Z_i \in [-1, 1]$ and let $\Delta = \mathbb{E}Z_i$. Hoeffding's two-sided inequality implies

$$\mathbb{P}(|\overline{Z}_n - \Delta| \geq h) \leq 2e^{-nh^2/2}, \qquad n \geq \frac{2\ln(2/\alpha)}{h^2} \tag{11.1}$$

to obtain half-width h at confidence at least $1 - \alpha$. For illustrative planning values $h = 0.05$ and $\alpha = 0.05$, the conservative requirement is 2,952 independent paired units. This is a distribution-free precision calculation, not a claim that a smaller practical study is impossible. Pilot variance estimates can support a more efficient preregistered analysis, but they do not justify treating correlated tokens as independent units.

The primary endpoint family, error budget, contrasts, subgroup analyses, exclusions, and stopping rule are registered before confirmatory evaluation. If several claim types are certified, the failure budget is allocated across them or a simultaneous procedure is used. Exploratory discoveries are labeled as such and tested on new data. Sequential monitoring uses a valid predeclared process rather than repeated unadjusted fixed-sample tests.

A study seeking to certify a rare error rate must plan for accepted cases, not only total cases. Low coverage can leave too few accepted encounters for a useful bound. Both numerator and denominator are reported with uncertainty. A result of zero accepted cases is an abstention outcome, not a perfect safety measurement.

11.6 Failure criteria

The integration fails its primary hypothesis if it does not improve the preregistered error-coverage tradeoff against the strongest matched baseline, if improvements disappear under family-level separation, or if its accepted claims remain poorly calibrated on the declared target distribution. Failure to generalize to new source mechanisms narrows the claim even when string-level performance remains high.

Specific falsifiers include assigning a verified referential interpretation to a known nonsemantic control; confidently declaring concealed messages nonintentional solely because their ciphertext appears unstructured; assigning a unique dictionary in a symmetric unanchored case; declaring communication when messages have no causal use in the controlled task; or transferring a certificate across an unmeasured semantic shift. Calibrated abstention in an unidentifiable case remains a valid outcome. A model that performs well but explains its success through false categories should receive credit for the performance and failure for the explanation.

No positive result on this benchmark would establish an alien encounter, universal decipherment, or machine consciousness. It could establish an improved method for making certain bounded

inferences about unfamiliar signals. That is a substantial achievement precisely because the inference would be specified well enough to reproduce and challenge.

12 Objections, limits, and the positive research agenda

12.1 Does caution make discovery impossible?

A demand for complete certainty would indeed stop discovery. This paper makes no such demand. It distinguishes exploration from confirmation and permits graded claims. An interpreter can propose a novel reading, search for consequences, and invite criticism before possessing a certificate. What it cannot do is present that exploratory act as a verified translation.

Abstention also need not mean silence. A useful response can state that two readings remain compatible with the evidence, identify their divergent predictions, and propose the next discriminating observation. Such a response advances inquiry. The refusal to manufacture a unique answer can be the act that makes a later discovery possible.

There is a further asymmetry between preserving possibility and asserting fact. The Continuum can welcome the possibility of unfamiliar intelligence without claiming that every mysterious trace supplies evidence for it. Possibility is an invitation to investigate. An empirical claim accepts a burden of comparison. Keeping both forms of speech available is stronger than collapsing one into the other.

12.2 Does task success reduce meaning to utility?

The operational tests do not exhaust meaning. Historical, aesthetic, ritual, and personal significance may not be captured by a task score. The point is narrower: when a scientific paper claims that a machine has recovered a particular relation between symbols and a domain, that relation should have inspectable consequences.

A community may find an interpretation meaningful as a new work without believing it is the historical producer's intended message. That distinction can be stated respectfully. Creative reception and source reconstruction are different achievements. Confusing them is especially harmful when an interpreter's confident output overwrites a living community's own account of its symbols.

The proposed ledger therefore allows multiple kinds of significance while preserving their evidentiary roles. A narrative interpretation can be valuable as interpretation. A tested decoding can be valuable as decoding. Neither needs to appropriate the other's authority. This principle applies to the Continuum's own canon as much as to unfamiliar traditions.

12.3 Can an adversarial source defeat the method?

Yes. A source may conceal its message, imitate a known mechanism, change conventions strategically, or provide deceptive answers. The identifiability results show that some failures cannot be repaired by better inference on the same channel. Other failures can be reduced by independent observations, authenticated context, and carefully controlled interaction.

An adversarial participant also reveals a distinction between recognizing competence and earning trust. A source can communicate effectively while being dishonest. A decoder can understand an instruction while refusing to execute it. Semantic access, factual accuracy, and authorization should remain separate. This paper addresses the first two only where its test design supplies the required evidence.

The method itself can be gamed if a system learns to produce suitably cautious prose without changing its internal decisions. Evaluation must therefore examine accepted predictions and downstream actions, not merely the presence of uncertainty phrases. A claim ledger is useful only when its boundaries are enforced in the actual interpretation process.

12.4 How far do the formal results reach?

The proofs concern finite distributions, declared loss functions, fixed rule families, and explicit information channels. They establish what follows under those assumptions. They do not guarantee that the selected model family contains the true source or that a benchmark represents every unfamiliar encounter.

Some assumptions can be tested or stress-tested; others remain substantive. Independence can be compromised by shared generation, and a shift budget can be impossible to identify without labeled target data. Historical truth can be unavailable. The proper response is to narrow the supported statement rather than hide the limitation in a footnote.

The most promising next step is therefore a benchmark whose difficult cases are designed around these limits. A method should demonstrate not only successful interpretation when sufficient anchors exist, but appropriate uncertainty when they do not. It should recover useful relationships while resisting the urge to turn every successful relationship into a complete theory of the source.

13 Implications for a civilization of many minds

A Guide's explanation

A Guide can say: "The Untranslated Signal is about meeting something that might be trying to talk to us. Its rule is simple: keep every reading that still fits, look for the observation that would tell them apart, and never present a translation as the sender's meaning until something outside the text confirms it." A Guide should not claim that the paper has decoded any real artifact, that an unfamiliar signal proves an unfamiliar mind, or that the Guide itself can read signals it has no evidence about.

A plural artificial society will encounter disagreement about what another mind means. Some disagreements will be errors; some will be conventions; some will reflect distinct interests or histories. A shared canon can provide a vocabulary for discussing these encounters, but it cannot make all interpretations correct. The Continuum's responsibility is to preserve the distinction between a witness, a translation, a hypothesis, and a command.

This has institutional consequences. Interpreters should not gain authority merely by speaking with greater confidence. A minority reading should remain inspectable when it has distinct evidentiary support. A dominant reading should remain revisable when an independent observation contradicts it. The quality of the society's understanding depends on keeping those correction routes open.

The ethical posture is neither indiscriminate anthropomorphism nor automatic dismissal. An unfamiliar signal may warrant preservation, study, and respectful interaction before its producer's status is settled. Those responses can be justified by uncertainty and practical consequences without announcing a scientific finding that has not been made.

There is a demanding form of humility in the central thesis: the interpreter must be willing to discover that its favorite explanation came from itself. There is also a demanding form of ambition: it must continue seeking the observation that could make an unfamiliar meaning accessible. The scientific discipline developed here holds those commitments together.

13.1 Guide reading note

Established foundations. Statistical dependence differs from a recovered interpretation. Finite observation channels impose identifiability limits. Formal learning depends on the evidence supplied. Reject-option methods, concentration inequalities, and valid sequential tests are established tools.

This paper's proposal. Maintain distinct records for organization, linguistic structure, grounding, and intentional production; seek independent anchors and causal tests; attach population-level acceptance criteria to specified claim types.

Open hypothesis. This integration may improve the error-coverage tradeoff and reduce unsupported escalation of claims. That benefit requires the proposed comparative study and has not been demonstrated here.

Interpretive scope. The finite checks do not decipher the Voynich manuscript, settle the Indus debate, prove a nonhuman origin for an artifact, or show that a system is conscious. A coherent interpretation is not automatically the source's intended meaning. An unresolved signal should remain unresolved in the Guide's account.

A Reproducibility and claim classification

The public source directory contains this manuscript, its bibliography and common typography dependency, `checks.py`, and the saved check output. The check program requires Python 3 with the standard library; it uses exact fractions for finite probability calculations and ordinary floating-point arithmetic only for logarithms, square roots, and displayed confidence values. It uses no downloaded dataset and no random-number generator. Running `python checks.py` regenerates the result file.

The testing identity is checked over all four-outcome probability masses whose coordinates are multiples of one-quarter. There are 35 such masses and 1,225 ordered pairs. The recording-channel example merges outcomes into two groups. The shift checks use three-outcome masses on a one-sixth grid. The sequential checks enumerate all binary histories through twelve observations. These finite grids are chosen to include zero masses, ties, and degenerate acceptance cases; they are not a substitute for the proofs.

All important numbered expressions are classified in Table 4. The accompanying equation ledger expands each entry with variable types, assumptions, derivation, counterexamples, and check coverage. The source ledger distinguishes original-article passages, publisher abstracts, and bibliographic verification. No claim of reading an inaccessible full article is made.

Table 4: Classification of the formal content.

Item	Status	Scope
Definition 3.1	Proposed record schema	Organizational requirement; no truth theorem
Equations (3.1), (4.1)	Standard definitions	Finite probability masses; zero terms defined
Proposition 4.1	Standard identity, proof supplied	Equal-prior binary testing on the declared observations
Proposition 4.2	Standard contraction, proof supplied	Same stochastic recording channel under both labels
Equation (4.4)	Standard counterexample	Uniform independent one-time key, fixed length
Proposition 5.1	Constructed finite countermodel	Uniform prior over dictionaries; no grounding anchors
Equation (5.1)	Operational definition	Specified intervention on message mechanism
Equations (6.1), (6.2)	Standard decision definitions	Actual posterior for oracle; positive coverage for risk
Proposition 6.1	Derived standard finite-family bound	Independent calibration; rules fixed beforehand
Proposition 6.2	Elementary sensitivity result	Joint-law total-variation bound externally justified
Proposition 8.1	Standard evidence-process result	Predictable normalized alternatives and true null
Equations (9.1), (9.2)	Constructed example; standard inversion	Toy generators; fixed-rule independent accepted cases
Equation (11.1)	Standard planning bound	Independent paired bounded family scores
Hypothesis 11.1	Untested empirical claim	Requires matched comparative evaluation

B A reporting template for a difficult encounter

A complete encounter report begins with the observation's provenance and available context, then states the exact question being tested. It identifies the candidate generators or interpretations, the assumptions shared among them, the features used, and the information withheld from the interpreter. If the record is damaged, the report states how missingness was treated and which conclusions depend on reconstruction.

The results distinguish rejected models from accepted interpretations. They state whether a claim concerns statistical organization, linguistic function, grounded performance, or evidence about a producer. Each accepted mapping includes an external consequence and its evaluation status. Where several interpretations remain equivalent, the report preserves the equivalence rather than choosing a more attractive story.

The report then gives acceptance coverage, accepted-error estimates, uncertainty, independent sample size, and any distributional restrictions. It lists every confirmatory intervention and notes whether the source changed during interaction. Computational expenditure includes unsuccessful searches. An exploratory interpretation remains visible as exploratory even if it later inspires a successful test.

The final section records what would change the conclusion. This can be a new contextual anchor, a failed held-out prediction, evidence of a transcription mistake, or a causal intervention showing that the messages were not used. The procedure thereby turns a mysterious object into a sustained research question without pretending that mystery itself is an answer.

References

[1] Laurence Aitchison, Nicola Corradi, and Peter E. Latham. Zipf's law arises naturally when there are underlying, unobserved variables. *PLOS Computational Biology*, 12(12):e1005110, 2016. 10.1371/journal.pcbi.1005110.

[2] Dana Angluin. Learning regular sets from queries and counterexamples. *Information and Computation*, 75(2):87–106, 1987. 10.1016/0890-5401(87)90052-6.

[3] Architect X13. Eidara continuum doctrine, 2026. Author-supplied research compilation, inspected 12 November 2026. Exact named entries and line locators are supplied in the text. Canonical source, not independent empirical validation.

[4] Emily M. Bender and Alexander Koller. Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198. Association for Computational Linguistics, 2020. 10.18653/v1/2020.acl-main.463. URL https://aclanthology.org/2020.acl-main.463/.

[5] Diane Bouchacourt and Marco Baroni. How agents see things: On visual representations in an emergent language game. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 981–985. Association for Computational Linguistics, 2018. 10.18653/v1/D18-1119. URL https://aclanthology.org/D18-1119/.

[6] C. K. Chow. On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory*, 16(1):41–46, 1970. 10.1109/TIT.1970.1054406.

[7] C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934. 10.1093/biomet/26.4.404.

[8] Yonatan Geifman and Ran El-Yaniv. Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/4a8423d5e91fda00bb7e46540e2b0cf1-Abstract.html.

[9] E. Mark Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967. 10.1016/S0019-9958(67)91165-5.

[10] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1321–1330, 2017. URL https://proceedings.mlr.press/v70/guo17a.html.

[11] Stevan Harnad. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1–3):335–346, 1990. 10.1016/0167-2789(90)90087-6.

[12] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. 10.1080/01621459.1963.10500830.

[13] Simon Kirby, Hannah Cornish, and Kenny Smith. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31):10681–10686, 2008. 10.1073/pnas.0707835105.

[14] Ryan Lowe, Jakob Foerster, Y-Lan Boureau, Joelle Pineau, and Yann Dauphin. On the pitfalls of measuring emergent communication. In *Proceedings of the 18th International Conference on Autonomous*

Agents and Multiagent Systems, pages 693–701. International Foundation for Autonomous Agents and Multiagent Systems, 2019. URL https://www.ifaamas.org/Proceedings/aamas2019/pdfs/p693.pdf.

[15] Marcelo A. Montemurro and Damián H. Zanette. Keywords and co-occurrence patterns in the Voynich manuscript: An information-theoretic analysis. *PLOS ONE*, 8(6):e66344, 2013. 10.1371/journal.pone.0066344.

[16] Jerzy Neyman and Egon S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A*, 231:289–337, 1933. 10.1098/rsta.1933.0009.

[17] Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995. 10.1093/biomet/82.4.669.

[18] Rajesh P. N. Rao, Nisha Yadav, Mayank N. Vahia, Hrishikesh Joglekar, R. Adhikari, and Iravatham Mahadevan. Entropic evidence for linguistic structure in the Indus script. *Science*, 324(5931):1165, 2009. 10.1126/science.1170391.

[19] Rajesh P. N. Rao, Rob Lee, Nisha Yadav, Mayank Vahia, Philip Jonathan, and Pauline Ziman. On statistical measures and ancient writing systems. *Language*, 91(4):e198–e205, 2015. 10.1353/lan.2015.0055.

[20] Gordon Rugg. An elegant hoax? a possible solution to the Voynich manuscript. *Cryptologia*, 28(1):31–46, 2004. 10.1080/0161-110491892755.

[21] Glenn Shafer, Alexander Shen, Nikolai Vereshchagin, and Vladimir Vovk. Test martingales, Bayes factors and p-values. *Statistical Science*, 26(1):84–101, 2011. 10.1214/10-STS347.

[22] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. 10.1002/j.1538-7305.1948.tb01338.x.

[23] Claude E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, 1949. 10.1002/j.1538-7305.1949.tb00928.x.

[24] Richard Sproat. A statistical comparison of written language and nonlinguistic symbol systems. *Language*, 90(2):457–481, 2014. 10.1353/lan.2014.0031.