Copied Consensus Is Not Objectivity

An Information-Theoretic Measure of Independent Recoverability

J. Farrow

Draft manuscript, revised v3 - May 2026

Abstract

Consensus can be produced either by many independent records of the same target or by many copied reports from a single source. Standard redundancy language often obscures this distinction. This paper defines effective objectivity as target-relative independent recoverability: the degree to which effectively independent fragments can form adequate readouts of a specified target. The proposed measure, FI_eff = max(0, log2(N_eff/k*)), combines an adequacy threshold k* with an effective independent fragment count N_eff estimated from target-conditioned residual correlations. The resulting Copy Gap, FI_naive - FI_eff, distinguishes genuine distributed recoverability from copied consensus. The framework is intentionally conservative. It does not measure truth by itself; it measures objectivity relative to a specified target. Simulations and boundary cases show that the measure increases under independent record formation, scales logarithmically with independent capacity, decays under drift, collapses under copied redundancy, and can expose false objectification when independent fragments converge on a false or captured target. The contribution is not the already-known claim that redundancy matters, but a portable operational diagnostic for separating independent recoverability from copied consensus across quantum, machine-learning, clinical-record, and social-informational systems.

Keywords: objectivity; redundancy; effective sample size; quantum Darwinism; copied consensus; documentation drift; representation learning; information theory

1. Introduction

A familiar failure of collective judgment is that agreement can be mistaken for evidence. A thousand reports may look like a thousand witnesses, but if they are copied from one source, they have little more evidential independence than the source itself. The problem is not merely social. It appears in sensor systems, model ensembles, clinical records, market narratives, and scientific measurement: a system may contain many fragments that appear to support the same claim, while informationally they amount to one repeated channel.

This paper proposes a compact diagnostic for that distinction. The central claim is not that objectivity is redundancy. That idea is already well represented in quantum Darwinism, distributed detection, partial information decomposition, and social epistemology. The narrower claim is that useful objectivity should be measured as effective independent recoverability: the number of effectively independent adequate readouts that can be formed for a specified target.

The measure is target-relative. Given a target X and fragments F1, ..., FN, the question is not whether the fragments are simply numerous or similar. The question is whether random subsets of them can recover X well enough, and whether those subsets are informationally independent rather than copies. This leads to a quantity measured in bits: FI_eff = max(0, log2(N_eff/k*)). Here k* is the smallest subset size that usually reaches adequacy, and N_eff is the effective number of independent fragments after conditioning out the shared target signal.

The paper is organized as follows. Section 2 places the proposal relative to known work. Section 3 defines the measure. Section 4 gives boundary cases. Section 5 derives the main equation from information-theoretic reasoning and states its approximation assumptions. Sections 6 and 7 summarize simulation and conceptual validation. Section 8 discusses applications, Section 9 states limitations, Section 10 gives falsifiable predictions, and Section 11 concludes.

2. What Is Already Known

Quantum Darwinism explains classical objectivity through redundant encoding of pointer-state information into environmental fragments. Recent functional-information formulations quantify objectivity as FI_QD(delta) = log2 R_delta, where R_delta counts environmental fragments that contain sufficient pointer information [1]. This paper takes that operational spirit seriously but shifts attention from environmental fragments alone to a general diagnostic of independent recoverability.

Statistics already contains the idea that correlated samples count as fewer independent samples. Effective sample size replaces a nominal count N with an effective count N_eff when dependence reduces estimation power [2]. The proposed measure imports that correction into objectivity measurement: many copied reports should not be treated as many independent records.

Partial information decomposition studies how multiple sources provide redundant, unique, or synergistic information about a target. Recent redundancy-bottleneck work formulates redundancy as an information-bottleneck problem [3]. The present measure is less general but more operational: it asks how many effectively independent adequate readouts can be formed from available fragments.

Machine-learning representation work, including the Platonic Representation Hypothesis, argues that independently trained models or modalities may converge toward shared representations of an underlying reality [4]. The present framework treats this as an empirical question. Representation similarity alone is not sufficient; the fragment must function as an adequate readout. Pilot tests found that raw embedding similarity can produce false-positive redundancy, while prediction-based readouts produce the expected low-to-high transition.

3. Formal Definition

Let X denote a specified target observable. Let F1, ..., FN denote fragments, records, sensors, model readouts, documentation surfaces, or channels that may contain information about X.

$$X \quad ; \quad \{F_1, \ldots, F_N\}$$

3.1 Adequate readouts

For a subset S of fragments, let I(X; F_S) be the mutual information between the target and the subset. A subset is delta-adequate if it contains at least a tolerated fraction of the available target information:

$$I(X; F_S) \geq (1 - \delta) I_{\max}(X)$$

In empirical settings, mutual information may be replaced by a calibrated readout score, such as prediction accuracy, log loss, semantic agreement, or human-rated adequacy. The critical requirement is that adequacy should measure functional recovery of the target, not mere geometric similarity among fragments.

3.2 The onset size k*

For a subset size k, define p_delta(k) as the probability that a randomly sampled k-fragment subset is adequate:

$$p_\delta(k) = P(I(X; F_S) \geq (1 - \delta) I_{\max}(X) \mid |S| = k)$$

Then define k* as the smallest subset size for which adequacy is typical at confidence level q:

$$k^*_{\delta, q} = \min\{k : p_\delta(k) \geq q\}$$

Small k* means many small subsets can recover the target. Large k* means the target is fragile, centralized, or only weakly distributed. If no subset is adequate, k* is undefined and the resulting objectivity score is zero.

3.3 Effective independence

Nominal fragment count N must be corrected for dependence. We estimate N_eff from the target-conditioned residual correlation matrix C. Conditioning on the target is essential: fragments should correlate through the target. What should be penalized is residual dependence after the shared target signal is removed.

$$N_{\text{eff}} = \frac{\left(\sum_i \lambda_i\right)^2}{\sum_i \lambda_i^2}$$

Here lambda_i are the eigenvalues of the residual correlation matrix. If fragments are approximately independent, N_eff is close to N. If the fragments are copies of one reporter, N_eff is close to 1.

3.4 Effective objectivity and the Copy Gap

$$FI_{\text{eff}} = \max\left(0, \log_2 \frac{N_{\text{eff}}}{k^*_{\delta, q}}\right)$$

The naive score uses N rather than N_eff:

$$FI_{\text{naive}} = \max\left(0, \log_2 \frac{N}{k^*_{\delta,q}}\right)$$

The Copy Gap is the difference between apparent redundancy and effective independent recoverability:

$$\text{Copy Gap} = FI_{\text{naive}} - FI_{\text{eff}}$$

A large Copy Gap indicates many fragments, but few effectively independent sources. This is the core diagnostic: copied consensus is separated from genuine distributed objectivity.

4. Boundary Behavior

Table 1 summarizes the intended semantics. The measure distinguishes private recoverability, redundant public objectivity, copied consensus, no-signal failure, and false objectification.

Scenario	k*	N_eff	FI_eff	Interpretation
One perfect fragment, no others	1	1	0	Perfect private recoverability, but zero redundant public objectivity.
100 perfect independent fragments	1	100	log2(100) = 6.64	Highly redundant objectivity.
100 copied fragments of one reporter	1	1	0	Copied consensus, not independent objectivity.
100 independent fragments requiring coalitions of 10	10	100	log2(10) = 3.32	Moderate objectivity through adequate coalitions.
No target information	undefined	any	0	No adequate readout exists.
False target independently reinforced	1	100	6.64 relative to false target	High target-relative objectification, not truth.

Table 1. Boundary behavior of FI_eff. The measure tracks redundant recoverability relative to a target, not truth by itself.

5. Derivation and Approximation Assumptions

The derivation begins with mutual information. A fragment subset S becomes an adequate observer when it contains enough information to recover the target within tolerance. The onset size k* is therefore the minimal resource cost, in fragments, required to form one adequate readout.

If N_eff effectively independent fragments are available and k* fragments are needed to form one adequate readout, then the system can support approximately N_eff/k* independent adequate readouts. This assumes that adequate readouts can be approximated by disjoint or weakly overlapping subsets of size k*, which is most defensible when fragments are exchangeable or near-exchangeable and k* is small relative to N_eff. When this approximation fails, the ratio should be treated as an operational estimate rather than an exact counting theorem.

Information theory measures multiplicity in bits by taking a base-2 logarithm. Therefore each doubling of the number of effectively independent adequate readouts adds approximately one bit of effective objectivity. The resulting measure is:

$$FI_{\text{eff}} = \max\left(0, \log_2 \frac{N_{\text{eff}}}{k^*_{\delta,q}}\right)$$

The clamp at zero ensures that systems unable to form even one adequate independent readout are not assigned negative objectivity. Negative values would indicate that the requested target is not effectively recoverable under the chosen tolerance and confidence settings.

6. Simulation Evidence

The simulations are not proof of a universal law. They validate whether the proposed measure behaves correctly under controlled conditions and whether it fails in the cases where it should fail. The most important tests are not merely positive cases, but negative and adversarial cases: no signal, copied reports, drift, false consensus, and threshold variation.

Test	Purpose	Key result	Interpretation
Quantum dephasing pilot	Reference against QD-style record formation	rise -> crossover -> plateau	Matches expected redundant-record dynamics.
ML prediction fragments	Independent small models as readouts	FI rose from 0 to plateau	Readout adequacy works better than embedding similarity.
Embedding controls	Check false positives from representation geometry	early false saturation	Representation similarity alone is too permissive.
Five synthetic systems	Cross-domain generative stress test	mean curve-collapse RMSE 0.073	Common low-to-high-to-plateau transition after rescaling.
No-signal control	Should produce no objectivity	FI_eff = 0	Passes null test.
Copied-consensus control	Many copies of one source	naive FI high, effective FI near 0	Copy Gap detects false redundancy.
Capacity scaling	Check log prediction	slope 0.980, R^2 = 0.9999	Doubling independent fragments adds about one bit.
Drift and false consensus	Test decay and target relativity	FI falls under drift; false target can objectify	Confirms target-relative semantics.

Table 2. Summary of simulation and stress-test results.

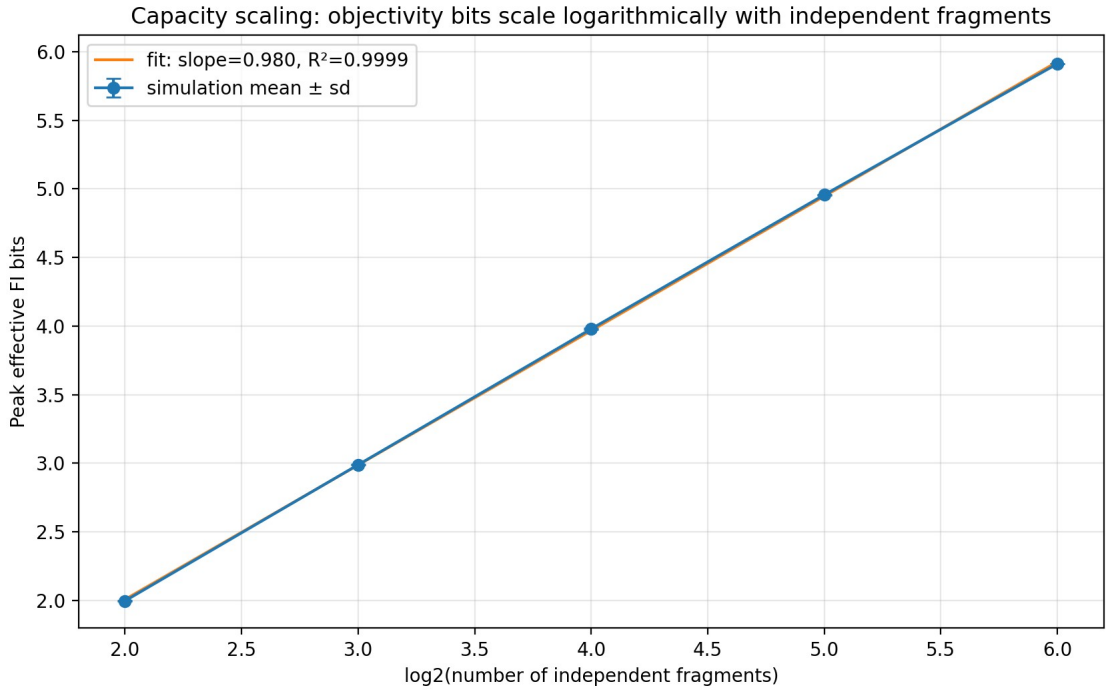

Figure 1. Capacity scaling. Peak effective objectivity scales approximately as log2(N) when independent fragments are individually adequate.

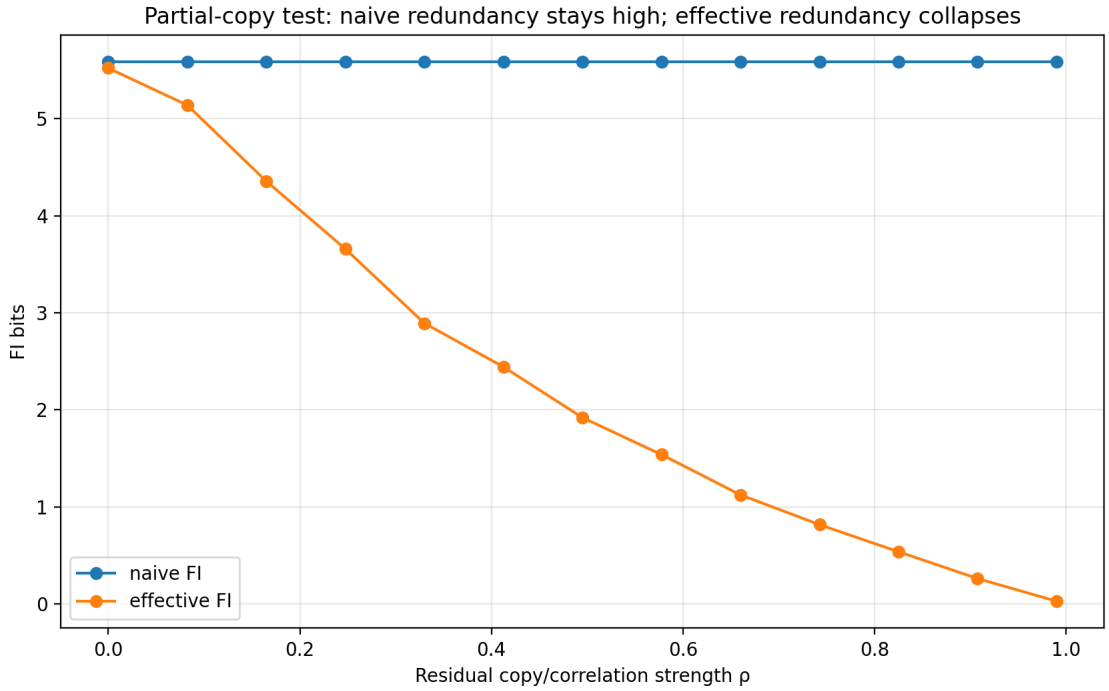

Figure 2. Partial-copy gradient. Naive FI remains high as copied dependence increases, while effective FI collapses toward zero.

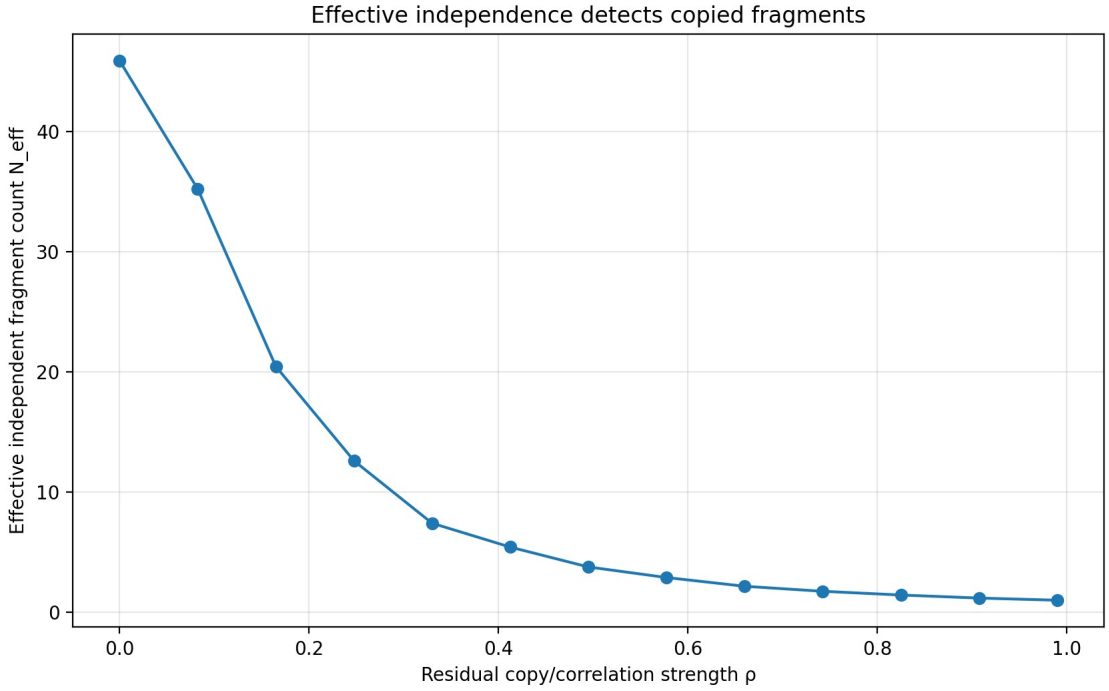

Figure 3. Effective independence correction. N_eff falls from many independent fragments toward one effective source as residual copying increases.

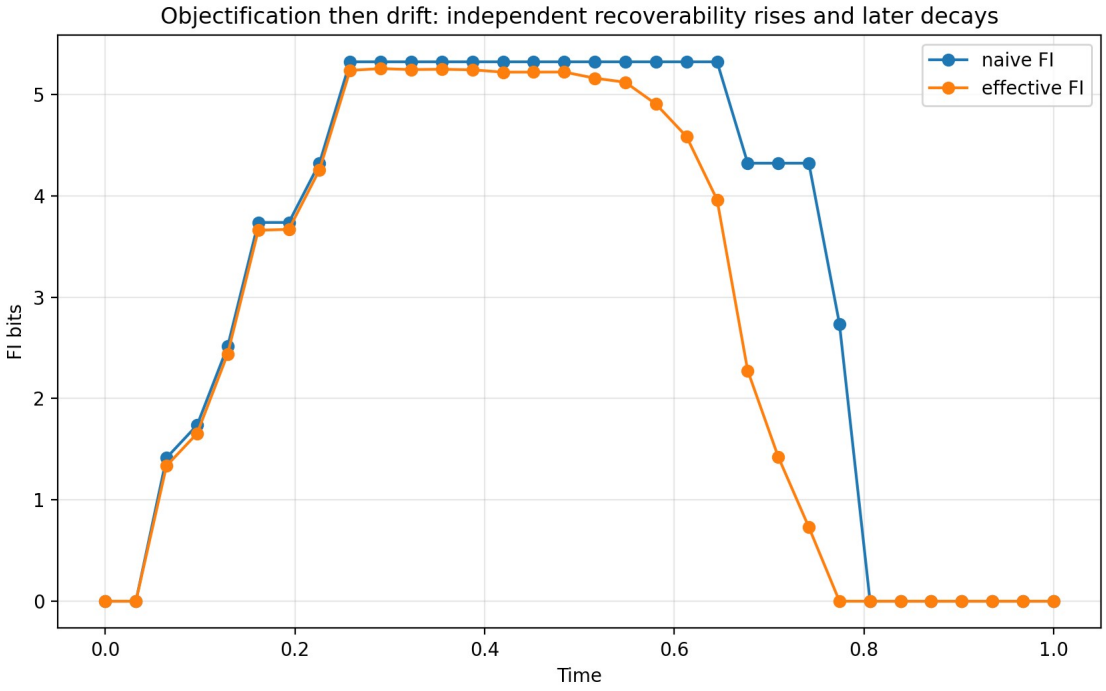

Figure 4. Objectification followed by drift. Independent recoverability rises, plateaus, and then decays as signal quality falls and reports become correlated.

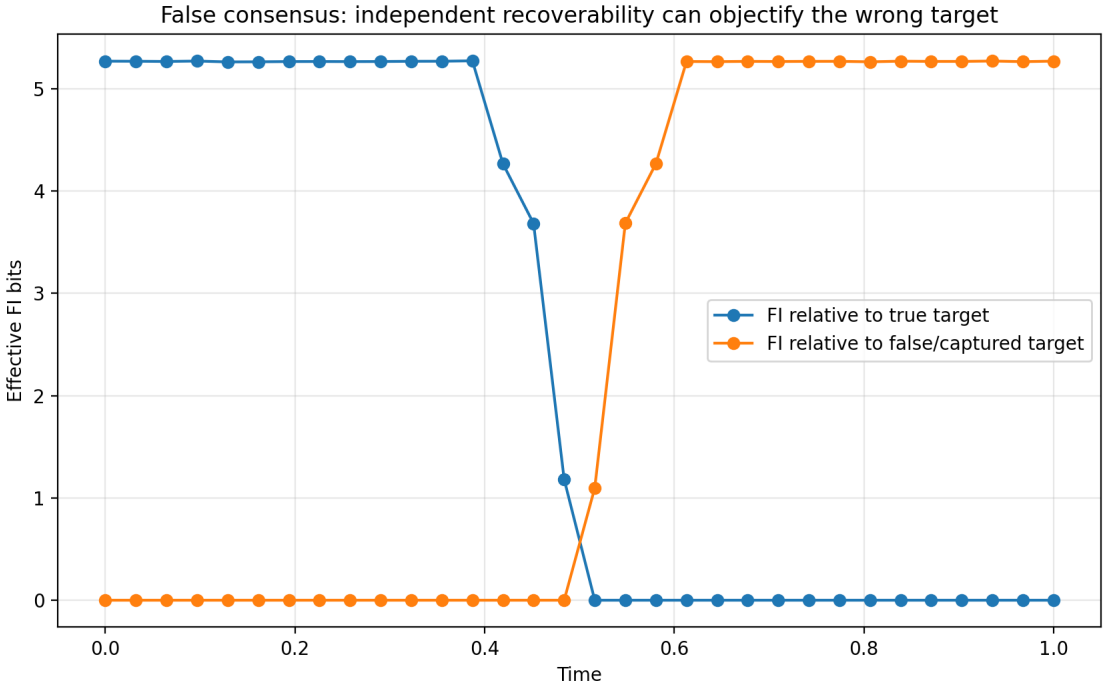

Figure 5. False consensus. Effective FI can rise for a false or captured target while falling for the true target, showing that the measure is target-relative.

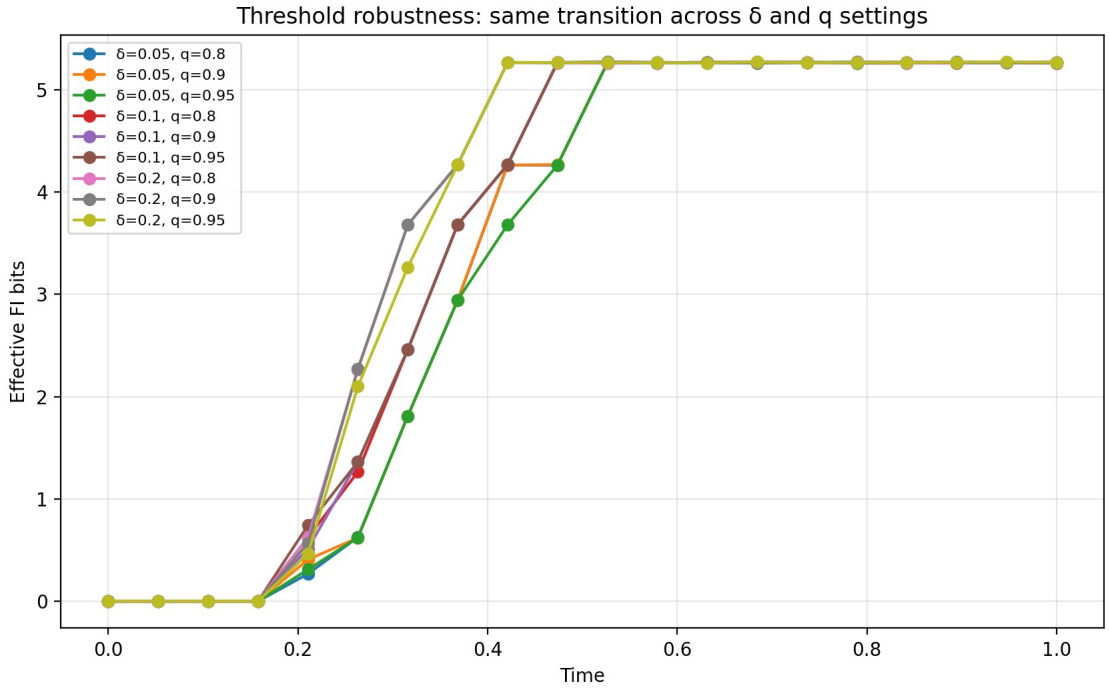

Figure 6. Threshold robustness. The same qualitative transition appears across tested delta and q settings.

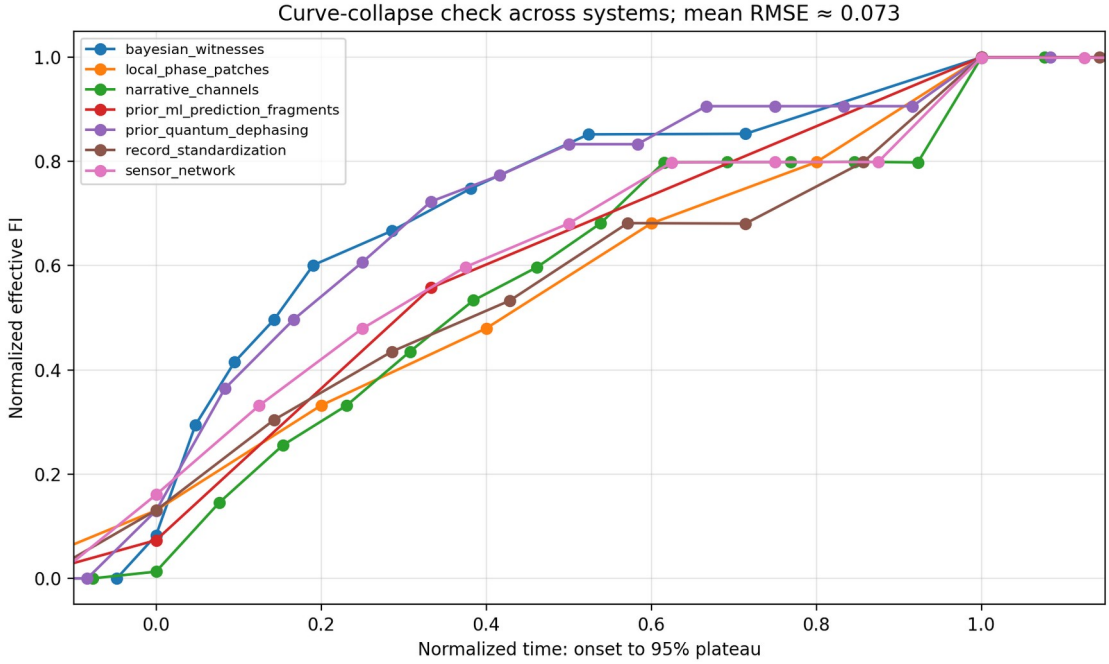

Figure 7. Curve-collapse check. Positive simulations show a common normalized low-to-high transition with mean RMSE about 0.073 and worst RMSE about 0.117.

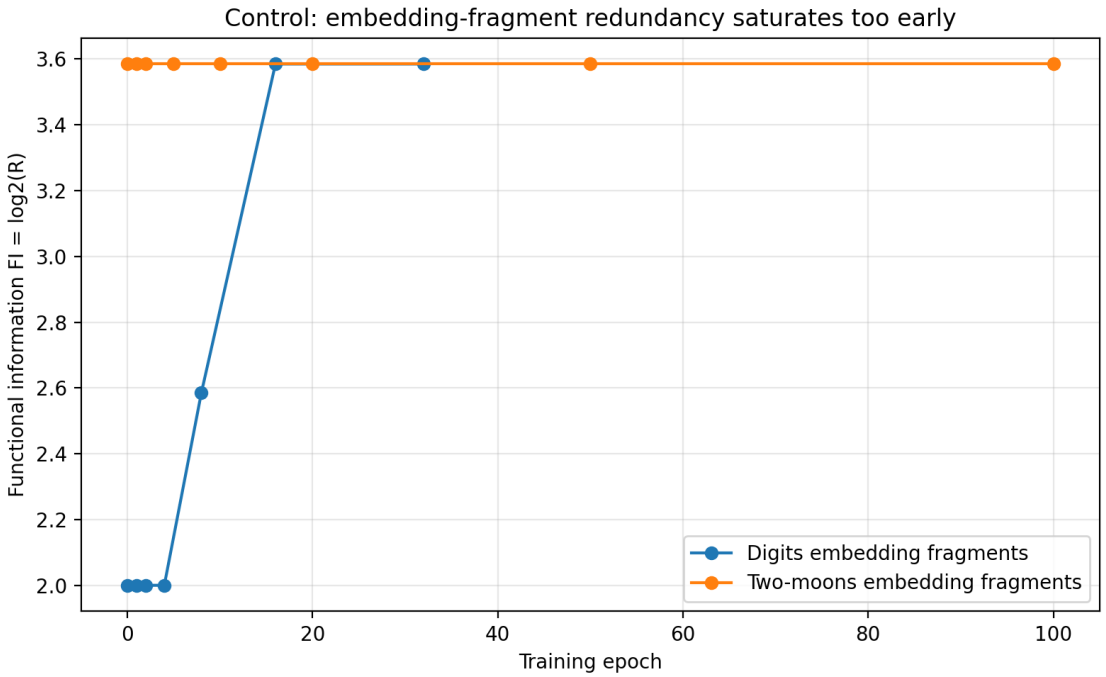

Figure 8. Embedding-control failure. Representation geometry alone produced early false-positive redundancy, supporting readout adequacy rather than similarity as the fragment criterion.

7. Conceptual Validation

The measure satisfies five internal consistency requirements. First, if many independent fragments are individually adequate, FI_eff is high. Second, if fragments are copied from one source, N_eff collapses and FI_eff is low even when naive N is large. Third, if no fragment subset carries target information, no k* exists and FI_eff is zero. Fourth, if independent adequate fragments double while k* remains fixed, FI_eff increases by one bit. Fifth, if record capacity saturates, FI_eff saturates because either k* cannot fall further or N_eff stops increasing.

The important conceptual distinction is between private accuracy and public objectivity. One perfect fragment can recover a target, but it does not provide redundancy. Thus FI_eff = 0 for N_eff = 1 and k* = 1 should not be read as ignorance. It means zero redundant objectivity. The measure is deliberately about public or intersubjective recoverability, not private certainty.

The copied-consensus case is the most important validation. Many copied reports can create high apparent agreement while contributing little independent information. The Copy Gap formalizes this difference: FI_naive can be high while FI_eff remains near zero. In social, clinical, ML, and sensor settings, that gap is often the phenomenon of interest.

8. Applications

8.1 Clinical and ABA documentation

In clinical documentation, fragments may include structured trial data, session notes, target graphs, caregiver reports, supervision comments, authorization language, and treatment-plan fields. A learner state is well documented when multiple independent surfaces can recover the same clinically relevant target. Falling FI_eff would indicate documentation drift: the clinical state is no longer redundantly recoverable from the record ecology.

This framing should be kept administrative rather than diagnostic. The measure can flag documentation health, consistency, and audit risk without making treatment decisions. This makes it suitable as a quality-control metric for clinical-record systems, especially if paired with human review.

8.2 Machine-learning ensembles

In model ensembles, fragments may be independently trained readouts. FI_eff can track whether learning has produced independent adequate recoverability of a target. Random-label controls should not show sustained FI_eff onset. A key methodological lesson from pilot tests is that embedding similarity alone is too permissive: random or weakly trained embeddings can preserve enough geometry to appear redundant. Prediction-based or task-adequate readouts better match the proposed definition.

This gives the measure a role in debates about representation convergence. The question becomes not whether representations look similar, but whether independent models become adequate observers of the same target and whether their recoverability remains independent after residual correction.

8.3 Narrative, market, and media systems

In social or market systems, fragments may include price action, options flows, news embeddings, social-media clusters, search interest, analyst reports, and policy signals. A narrative becomes objectified when many independent channels recover the same latent regime. But if all channels copy the same source, FI_naive may be high while FI_eff remains low. That pattern is precisely copied consensus.

The false-consensus test clarifies the ethical risk. A false target can become highly objectified if independent channels learn to recover it. Therefore FI_eff should not be used as a truth oracle. It should be paired with causal grounding, adversarial reference channels, and target validation.

9. Limitations

First, target selection is not solved by the measure. FI_eff is always relative to a specified target. If the target is false, corrupted, or poorly defined, the measure can quantify the objectification of error. Truth requires additional grounding.

Second, target-conditioned independence is easiest when the target is known. In real systems, the target may be latent, noisy, or inferred from the same fragments being evaluated. In such cases, residual conditioning can under-correct or over-correct dependence.

Third, the measure is conservative for highly synergistic systems. If only the whole set of fragments can recover the target, then k* is near N_eff and FI_eff is low. Such systems may contain distributed information, but not redundant public recoverability in the sense defined here. If many independent synergistic coalitions exist, FI_eff can still be positive; the limitation applies most strongly when adequate readout requires nearly all fragments.

Fourth, the N_eff estimator used here is one practical choice, not a final theory of independence. Bayesian effective sample size, graph-based source modeling, causal discovery, or PID-based redundancy estimates may be superior in some domains.

Fifth, the simulations are synthetic. They validate edge behavior and stress-test logic, but they do not establish empirical predictive utility in clinical, ML, or social systems. Real-data validation remains the main next step.

10. Falsifiable Predictions

Prediction	Domain	Expected result	Failure condition
Copy-gap test	Media / records / markets	Copied channels show high FI_naive but low FI_eff.	Effective FI remains high despite known copying.
Learning test	ML ensembles	Valid labels produce FI_eff onset; random labels do not.	Random labels show comparable sustained FI_eff.
Drift test	Clinical documentation	Falling FI_eff predicts documentation inconsistency or audit risk.	FI_eff does not precede or track drift.
Capacity test	Sensor/model systems	Doubling independent adequate fragments adds about one bit.	Peak FI does not scale with log2(N_eff).

Table 3. Falsifiable predictions and failure conditions.

11. Conclusion

Objectivity is not merely agreement, and consensus is not necessarily evidence. A system may contain many reports and still possess little independent recoverability if those reports are copied, correlated, or captured. This paper formalizes that distinction with an information-theoretic order parameter: FI_eff = max(0, log2(N_eff/k*)).

The measure is intentionally modest. It does not prove truth. It measures the abundance of effectively independent adequate readouts of a specified target. Its value lies in distinguishing public redundancy from private accuracy, copied consensus from independent objectivity, and drift from stable recoverability.

The next step is empirical. If FI_eff predicts clinical documentation drift, detects copy-driven hype, separates random-label learning from real target recovery, or improves audit of distributed record systems, then it becomes more than a useful definition. It becomes an operational instrument for measuring when a target has become independently recoverable in the world.

Appendix A. Minimal Operational Algorithm

Specify the target X and the fragment set F1, ..., FN.

Choose an adequacy score and tolerance delta.

For each subset size k, sample many k-fragment subsets and compute adequacy probability p_delta(k).

Set k* to the smallest k where p_delta(k) >= q.

Estimate N_eff from target-conditioned residual dependence among fragments.

Compute FI_eff = max(0, log2(N_eff/k*)).

Compute Copy Gap = FI_naive - FI_eff, where FI_naive uses nominal N.

Interpret high FI_eff as independent recoverability relative to X; interpret high Copy Gap as copied or dependent consensus.

References

[1] Tank, A. B. (2025). Functional Information in Quantum Darwinism: An Operational Measure of Objectivity. arXiv:2509.17775.

[2] Stan Development Team. Stan Reference Manual: Effective Sample Size. The manual describes N_eff as the number of independent samples with the same estimation power as autocorrelated samples.

[3] Kolchinsky, A. (2024). Partial Information Decomposition: Redundancy as Information Bottleneck. Entropy, 26(7), 546.

[4] Huh, M., Cheung, B., Wang, T., & Isola, P. (2024). The Platonic Representation Hypothesis. arXiv:2405.07987.

[5] Zurek, W. H. (2009). Quantum Darwinism. Nature Physics, 5, 181-188.

[6] Williams, P. L., & Beer, R. D. (2010). Nonnegative decomposition of multivariate information. arXiv:1004.2515.