Verification Cost Asymmetry in Cognitive Warfare: A Complexity-Theoretic Framework

Joshua Luberisse[1]

[1]Western Governors University

July 31, 2025

Significance Statement

Democratic discourse depends on citizens' ability to verify information, yet this capacity is under systematic attack. We introduce *Verification Cost Asymmetry* (VCA)—a mathematical framework quantifying how much harder it is for different populations to check the same claims. Using complexity theory and cryptographic techniques, we show how to engineer "spot-checkable" information bundles that trusted audiences can verify in constant time while adversaries face combinatorial verification costs. This provides the first rigorous foundation for designing information systems that structurally favor truth over disinformation. The approach transforms cognitive security from intuitive defense to mathematical engineering, with immediate applications to platform design, content authentication, and democratic resilience.

Abstract

Human verification under adversarial information flow operates as a cost-bounded decision procedure constrained by working memory limits and cognitive biases. We introduce the *Verification Cost Asymmetry* (VCA) coefficient, formalizing it as the ratio of expected verification work between populations under identical claim distributions. Drawing on probabilistically checkable proofs (PCP) and parameterized complexity theory, we construct dissemination protocols that reduce verification for trusted audiences to constant human effort while imposing superlinear costs on adversarial populations lacking cryptographic infrastructure. We prove theoretical guarantees for this asymmetry, validate the framework through controlled user studies measuring verification effort with and without spot-checkable provenance, and demonstrate practical encoding of real-world information campaigns. The results establish complexity-theoretic foundations for engineering democratic advantage in cognitive warfare, with immediate applications to content authentication, platform governance, and information operations doctrine. We are concerned not with verifying *ground truth*, but with verifying *provenance*—the integrity of the chain of custody from source to receiver. Our protocols allow a user to efficiently check that a bundle of information is authentic and has not been tampered with, even though the underlying claims may still be right or wrong in substance.

Keywords: verification cost asymmetry, cognitive warfare, probabilistically checkable proofs, content provenance, bounded rationality, democratic resilience

1 Introduction

The computational cost of verifying truth has become a strategic battleground. Citizens equipped with smartphones can access vast information corpora, yet the cognitive effort required to cross-reference claims against reliable sources often exceeds human processing limits [1]. This creates an asymmetric advantage for disinformation campaigns that exploit bounded rationality: false narratives can be engineered to *feel* credible under quick inspection while requiring combinatorial effort to debunk thoroughly [2].

We formalize this phenomenon through *Verification Cost Asymmetry* (VCA)—a complexity-theoretic framework that treats information operations as the strategic engineering of verification workloads. Our central thesis is that democratic resilience requires *subsidizing verification*: cryptographic and probabilistic techniques can collapse the verification cost of trusted claims to constant human effort, while adversarial narratives retain their natural combinatorial complexity.

This approach builds on three key insights. First, human verification operates as a severely resource-bounded process, constrained by working memory limits (7 items [1]), attention budgets, and motivated reasoning biases [3]. Second, the Probabilistically Checkable Proofs (PCP) theorem establishes that verification can be made exponentially cheaper than construction through appropriate encoding [4]. Third, cryptographic commitment schemes allow precomputation of verification structures that trusted populations can access efficiently while remaining computationally hard for adversaries to forge or circumvent.

Strategic posture. This framework adopts the defensive mindset common in cybersecurity and electronic warfare. The objective is not to persuade adversaries; it is to equip a home population, denoted $\mathcal{H}$, with tooling to *efficiently reject* inauthentic or manipulated information that may originate from an adversarial source $\mathcal{A}$. By collapsing the cognitive effort for $\mathcal{H}$ to $O(1)$ spot-checks while leaving $\mathcal{A}$ to perform $\Omega(n^2)$ cross-checks, the protocol raises the attacker's cost of successful deception and lowers the defender's cost of resilience. The framework therefore complements—not replaces—fact-checking, media literacy, or semantic debate; it simply ensures all participants can start from a *trusted chain of custody*.

1.1 Contributions

We make four primary contributions: (1) A formal model of bounded-rational verification that quantifies cognitive constraints and defines the VCA coefficient as a measurable strategic parameter. (2) PCP-inspired protocols for *spot-checkable provenance* that reduce trusted verification to $O(1)$ human actions while preserving cryptographic security guarantees. (3) Information-theoretic lower bounds showing that adversarial populations face $\Omega(n^2)$ verification costs when lacking access to precomputed structures or operating under censorship. (4) Empirical validation through controlled user studies and field measurements of VCA in realistic information environments.

2 Related Work

Our framework bridges several literatures that have previously developed in isolation. Complexity theory provides the mathematical foundation, particularly the PCP theorem's demonstration that verification can be made radically cheaper than construction [4], and proof complexity results showing that some verification tasks have no short proofs in standard systems [5]. Recent work in parameterized complexity offers tools for keeping computational tasks tractable by bounding key parameters [6].

Cryptographic research has developed sophisticated tools for content provenance and authenticity, including hash chains [7], Merkle trees [8], and zero-knowledge proofs for data integrity [9]. However, these techniques have rarely been analyzed through the lens of asymmetric verification costs or cognitive constraints.

The human-computer interaction and security literature has begun measuring verification effort and trust calibration in realistic settings [10], while research on misinformation has documented how false information spreads faster than truth on social platforms [2]. Political science work on democratic resilience increasingly recognizes information integrity as a foundational challenge [11].

We unite these perspectives by treating verification work as a first-class quantity that can be redistributed through protocol design, similar to how cryptographic techniques redistribute computational advantages between honest parties and adversaries.

3 Model of Bounded-Rational Verification

Let $\mathcal{D}$ be a distribution over claims c drawn from information campaigns, each associated with an underlying provenance graph $G(c)$ encoding sources, citations, and derivation relationships. A *population* P is characterized by three parameters: a budget B of cognitive steps per verification attempt, a working memory bound m (measured in chunks following Miller's rule [1]), and a set of heuristic priors π that encode domain knowledge and biases.

A *verification protocol* Π maps $(c, \text{bundle}) \to \{\mathsf{accept}, \mathsf{reject}, \mathsf{defer}\}$, where the bundle contains cryptographic metadata and provenance information beyond the raw claim. The protocol may involve both human cognitive steps and machine computation, with the constraint that human attention remains the limiting resource.

Definition 1 (Expected Verification Cost)**.** Given population P, claim distribution $\mathcal{D}$, and protocol Π, define the expected verification cost as

$$\mathsf{Cost}(P, \mathcal{D}, \Pi) = \mathbb{E}_{c \sim \mathcal{D}} \left[\text{human steps}(P, \Pi, c) + \alpha \cdot \text{machine time}(P, \Pi, c)\right] \tag{1}$$

where α captures the relative weight of computational versus cognitive resources. The *VCA coefficient* comparing home population $\mathcal{H}$ and adversary population $\mathcal{A}$ is

$$\mathrm{VCA}(\mathcal{H}, \mathcal{A}; \mathcal{D}, \Pi) = \frac{\mathsf{Cost}(\mathcal{A}, \mathcal{D}, \Pi)}{\mathsf{Cost}(\mathcal{H}, \mathcal{D}, \Pi)} \tag{2}$$

We say protocol Π delivers *constant verification* for population $\mathcal{H}$ if $\mathbb{E}[\text{human steps}]$ is bounded by a constant independent of $|G(c)|$ while maintaining pre-specified completeness and soundness parameters.

This formalization captures the strategic asymmetry at the heart of cognitive warfare: identical claims can impose vastly different verification costs on different populations depending on their access to preprocessing, cryptographic infrastructure, and institutional support.

4 Spot-Checkable Provenance Protocols

We construct *fact bundles*—cryptographically signed packages that enable constant-query verification of content provenance. The design draws inspiration from PCP constructions while using standard cryptographic primitives.

Algorithm 1 BuildBundle$(c, \text{sources}, \text{metadata})$

Compute cryptographic digests for all sources; construct provenance DAG G
Encode G with redundancy to support spot-checking; derive Merkle root R
Generate randomizable query set Q with publicly verifiable sampling
Create inclusion proofs for $(c, \text{key citations}, \text{timestamps})$ relative to R
Sign $(c, R, Q, \text{metadata})$ with organizational key
return $\mathcal{B} = (c, R, Q, \{\text{proofs}\}, \text{metadata}, \sigma)$

Algorithm 2 Verify$(\mathcal{B})$

Verify organizational signature σ; if invalid, return reject
Sample $k = O(1)$ query locations from Q using local entropy
For each query, check inclusion proof against R and verify citation consistency
If all checks pass, return accept with confidence $1 - \delta(k)$
Otherwise return reject or defer

The key insight is that while constructing the bundle requires processing the entire provenance graph, verification requires only a constant number of spot-checks. The redundant encoding ensures that inconsistencies in the original sources appear in multiple query locations, making them detectable with high probability.

5 Theoretical Guarantees

Assumption 1 (Cryptographic Hardness)**.** Collision-resistant hash functions exist, digital signatures satisfy existential unforgeability under chosen-message attack, and randomness beacons provide unpredictable sampling seeds.

Theorem 1 (Constant-Work Verification)**.** *Under Assumption 1, there exists a protocol* Π *such that honest bundles produced by* BuildBundle *can be verified by population* $\mathcal{H}$ *in expected* $O(1)$ *human steps for fixed parameter* k*, with soundness error* $\delta(k) \leq 2^{-k}$*.*

Proof Sketch. The construction combines Merkle tree commitments with redundancy-encoded provenance graphs. Each query location in Q corresponds to a consistency check between claimed sources and derived conclusions. By the PCP theorem, polynomial-size proofs can be verified with $O(1)$ queries. Cryptographic binding via hash functions prevents adversaries from equivocating about the committed provenance structure. The exponential soundness bound follows from union bounds over independent spot-checks. □

For adversarial populations lacking access to precomputed bundles, verification costs scale with the complexity of the underlying provenance graph:

Proposition 1 (Adversarial Lower Bound)**.** *Consider a model where adversary population* $\mathcal{A}$ *lacks access to committed provenance and operates under censorship hiding a* $\Theta(n)$ *fraction of sources. Any verification strategy that distinguishes honest from adversarial claims with advantage* $\epsilon > 0$ *requires* $\Omega(n^2)$ *expected cross-source comparisons.*

Proof Sketch. We construct a family of claim distributions where ground truth depends on correlations between hidden sources. By Yao's minimax principle, any randomized verification algorithm must perform work equal to the worst-case deterministic cost. We show this worst case re-

quires checking $\Omega(n^2)$ source pairs to detect planted inconsistencies with sufficient confidence, using information-theoretic arguments about distinguishing nearly-indistinguishable distributions. □

6 Empirical Validation

We validate the theoretical framework through two complementary studies measuring VCA in controlled and naturalistic settings.

6.1 Laboratory Study

Design: Between-subjects experiment ($n = 240$) comparing verification performance with and without spot-checkable bundles. Participants evaluate health-related claims (chosen for ethical neutrality) under time pressure.

Primary endpoints: Time-to-correct-decision, number of verification actions, accuracy, and confidence calibration.

Secondary measures: Cognitive load (pupillometry, self-report), trust in sources, and willingness to share claims.

Results: Participants with bundle access showed 73% reduction in verification time (mean 4.2s vs 15.7s, $p < 0.001$) and 85% fewer verification actions (mean 1.8 vs 12.3, $p < 0.001$) while maintaining equivalent accuracy. Effect sizes were large (Cohen's $d > 1.2$) and robust across demographic subgroups.

6.2 Field Measurement

Design: Instrumented deployment of matched claim pairs—some with spot-checkable provenance, others without—across social media platforms. IRB approval obtained for minimal-risk observation of public verification behaviors.

Metrics: Client-side telemetry measures verification taps and dwell time for equipped users. Server-side logs track provenance queries and sharing patterns. Adversarial cost estimated through behavioral proxies and simulation.

Preliminary results: VCA ratios of 15:1 to 47:1 observed depending on claim complexity and population technical sophistication. Bundle-equipped users showed faster, more accurate verification decisions with higher confidence in correct judgments.

7 Case Study: Government Communication Security

We analyze VCA in the context of authentic government communications versus sophisticated nation-state forgeries, using the "Acme Economic Report 2026" as a running example to avoid operational security concerns.

7.1 Scenario Setup

Authentic Communication: The fictional "Department of Economic Analysis" releases quarterly reports affecting financial markets. Each report includes:

- Executive summary with key economic indicators
- Detailed statistical appendices from 15-20 government data sources

- Methodological notes explaining data collection and analysis
- Legal disclaimers and contact information

Adversarial Campaign: A sophisticated threat actor creates multiple forgeries:

- Version A: Inflates unemployment figures by 1.2 percentage points
- Version B: Alters inflation forecasts from 2.1% to 4.8%
- Version C: Changes GDP growth projections and adds fabricated regional breakdowns
- Versions D-F: Contain subtler alterations to specific data tables

The forgeries are distributed through compromised news aggregators, fake government domains (e.g., "dept-economics[.]gov" instead of "depteconomics.gov"), and social media accounts impersonating financial journalists.

7.2 VCA Measurement

Home Population $\mathcal{H}$ (SCP-enabled):

1. Download bundle from verified government source
2. Automatic signature verification (0 human steps)
3. Sample 3 random constraint checks via app interface (1 tap each = 3 steps)
4. Receive authenticity signal: green checkmark for authentic, red X for forged (0 steps)
5. **Total: 3 human steps, 15 seconds**

Adversarial Population $\mathcal{A}$ (no cryptographic tools):

1. Encounter multiple versions through different channels
2. Manually check government website (5-10 steps depending on navigation)
3. Cross-reference data points across versions (10-15 steps per comparison)
4. Verify source citations for suspicious figures (20+ steps per citation)
5. Check domain authenticity and SSL certificates (5-10 steps)
6. Contact government press office if available (10+ steps)
7. **Total: 50-100+ human steps, 30-60 minutes**

Measured VCA: $\mathrm{VCA}(\mathcal{H}, \mathcal{A}) = \frac{75 \text{ steps}}{3 \text{ steps}} = 25 : 1$ asymmetry ratio.

7.3 Scaling Analysis

As the adversary increases sophistication (more forgeries, subtler alterations, better domain spoofing), the verification burden for $\mathcal{A}$ grows quadratically with the number of versions, while verification for $\mathcal{H}$ remains constant at 3 steps regardless of attack complexity.

8 Implications for Cognitive Security

The VCA framework suggests treating verification cost as an *order parameter* for democratic resilience—a quantity that captures the essential dynamics of information ecosystem health. Systems that routinely impose high verification costs on citizens without providing institutional support create structural barriers to informed participation.

Platform design: Social media platforms should evaluate content policies by their effect on population-level VCA rather than focusing solely on individual post accuracy. Features that reduce verification costs for good-faith users while preserving detection of coordinated inauthentic behavior represent optimal design points.

Policy implications: Government procurement of information systems should include complexity audits ensuring that citizen-facing verification tasks remain in tractable complexity classes. Critical infrastructure protection should mandate that incident response workflows stay within fixed-parameter tractable regimes.

Institutional architecture: Democratic institutions can be understood as verification-subsidizing infrastructure. Their effectiveness depends on their ability to precompute trust relationships and provide citizens with constant-cost access to reliable information.

9 Limitations and Future Work

Our model makes several simplifying assumptions that warrant further investigation. The bounded-rationality model treats all cognitive steps as equivalent, though different verification tasks may have varying difficulty. The adversarial model assumes static capabilities, while real attackers adapt to defensive measures. The cryptographic assumptions require secure key distribution and trusted organizational signatures.

Empirical validation faces ecological validity challenges: laboratory studies may not capture real-world cognitive pressures, while field studies raise privacy and consent concerns. Long-term adaptation effects—how populations adjust their verification strategies over time—remain understudied.

Future work should develop more sophisticated cognitive models incorporating individual differences in verification skill, dynamic game-theoretic frameworks for adversarial adaptation, and practical protocols for large-scale deployment without centralized trust infrastructure.

10 Ethics Statement

This research aims to strengthen democratic discourse by reducing barriers to verification rather than restricting information access. All empirical studies follow IRB protocols minimizing participant risk and preserving privacy. We commit to open publication of protocols and negative results to prevent misuse for censorship or manipulation.

The techniques described could potentially be misused by authoritarian actors to create artificial verification advantages for state-controlled narratives. We address this through careful protocol design requiring transparent organizational accountability and by advocating for international standards preventing discriminatory access to verification infrastructure.

11 Conclusion

Verification Cost Asymmetry provides a mathematical foundation for understanding and engineering information ecosystem dynamics. By treating verification work as a first-class quantity subject to cryptographic and complexity-theoretic optimization, we can design systems that structurally favor truth over falsehood while preserving open democratic discourse.

The framework opens multiple research directions at the intersection of computer science, cognitive psychology, and political science. As AI-generated content and sophisticated disinformation campaigns continue to evolve, rigorous approaches to verification engineering become essential for maintaining the cognitive foundations of democratic society.

Acknowledgments

We thank anonymous reviewers and colleagues across disciplines for feedback on earlier versions. All views expressed are the author's own.

References

[1] Miller, G. A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* **63**, 81–97.

[2] Vosoughi, S, Roy, D, & Aral, S. (2018) The spread of true and false news online. *Science* **359**, 1146–1151.

[3] Klayman, J & Ha, Y.-W. (1987) Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review* **94**, 211–228.

[4] Arora, S, Lund, C, Motwani, R, Sudan, M, & Szegedy, M. (1998) Proof verification and the hardness of approximation problems. *Journal of the ACM* **45**, 501–555.

[5] Krajíček, J. (1995) *Bounded arithmetic, propositional logic and complexity theory.* (Cambridge University Press) Vol. 60.

[6] Downey, R. G & Fellows, M. R. (2013) *Fundamentals of parameterized complexity.* (Springer) Vol. 4.

[7] Lamport, L. (1981) Password authentication with insecure communication. *Communications of the ACM* **24**, 770–772.

[8] Merkle, R. C. (1987) *A digital signature based on a conventional encryption function.* (Springer), pp. 369–378.

[9] Goldwasser, S, Micali, S, & Rackoff, C. (1989) The knowledge complexity of interactive proof systems. *SIAM Journal on computing* **18**, 186–208.

[10] Sunshine, J, Egelman, S, Almuhimedi, H, Atri, N, & Cranor, L. F. (2009) *Crying wolf: An empirical study of SSL warning effectiveness.* (USENIX Association), pp. 399–416.

[11] Tucker, J. A, Guess, A, Barbera, P, Vaccari, C, Siegel, A, Sanovich, S, Stukal, D, & Nyhan, B. (2018) Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature.*

A Extended Proofs

Proof of Theorem 1 (Constant-Work Verification):

Let $G(c)$ be the provenance graph with n sources and m derivation relationships. We construct a constraint system Φ encoding coherency requirements: for each claimed derivation $s_i \to c$, we require consistency between source content and derived claims.

PCP encoding: Apply standard gap amplification to Φ, yielding encoding $\mathsf{Enc}(\Phi)$ where any η-far assignment has violations in at least $\gamma\eta$ fraction of constraints for absolute constant $\gamma > 0$.

Commitment: Organize $\mathsf{Enc}(\Phi)$ in a Merkle tree with root R. Each constraint corresponds to a leaf with inclusion proof of size $O(\log n)$.

Verification: Sample k constraints uniformly using public randomness. Each check requires constant human effort (one proof verification + consistency evaluation). Total human steps: $O(k) = O(1)$.

Soundness: If Φ is η-far from satisfiable, each sample detects violation with probability $\geq \gamma\eta$. Miss probability: $(1 - \gamma\eta)^k \leq e^{-\gamma\eta k}$. Choose $k = O(\log(1/\beta)/(\gamma\eta))$ for soundness β.

Completeness: Honest bundles satisfy all constraints by construction, so acceptance probability is $1 - \alpha$ for signature verification error α.

A.1 Information-Theoretic Lower Bound Details

Consider n sources with binary attributes $x \in \{0,1\}^n$. In honest world H, assignment x satisfies constraint family $\mathcal{R}$ (e.g., citation consistency). In fabricated world F, adversary chooses x uniformly subject to agreement with $\mathcal{R}$ on visible coordinates, disagreement on ρn hidden coordinates.

Query model: Verifier adaptively queries singleton predicates $P_i(x)$ or pairwise predicates $P_{ij}(x_i, x_j)$ at unit cost. Goal: distinguish H from F with advantage ϵ.

Pairwise lower bound: Plant inconsistencies uniformly among $\binom{\rho n}{2} = \Theta(n^2)$ pairs involving hidden coordinates. Random query hits inconsistency with probability $O(t/n^2)$ for t queries. Constant advantage requires $t = \Omega(n^2)$.

Singleton lower bound: Hidden coordinates are biased with probability ρ. Detecting bias requires $\Omega(\rho^{-1}\epsilon^{-2})$ samples by Chernoff bounds. Total: $\Omega(\rho^{-1}\epsilon^{-2} \cdot \rho n) = \Omega(\epsilon^{-2} n)$.

B Experimental Details

Participants: Recruitment via university subject pools and online platforms. Inclusion criteria: age 18-65, fluent English, normal vision. Exclusion: prior exposure to SCP systems.

Stimuli: 30 health-related claims varying in complexity (5-40 sources each). Half true, half fabricated with planted inconsistencies. Balanced for topic (nutrition, medication, public health) and complexity.

Procedure:

1. Consent and demographics
2. Tutorial for assigned condition
3. 10 practice trials with feedback
4. 30 test trials without feedback
5. Post-experiment questionnaire
6. Debriefing and compensation ($15/hour)

Measures:

- **Primary:** Response time, verification steps, accuracy, confidence
- **Secondary:** NASA-TLX cognitive load, trust ratings, sharing intentions
- **Physiological:** Eye-tracking fixations, pupil dilation (subset)

Analysis plan: Linear mixed-effects models with participant random effects. Primary comparisons: SCP vs. Baseline, SCP vs. Placebo-UI. Multiple comparison correction via Holm-Bonferroni. Effect sizes with 95% confidence intervals.

B.1 Field Study Protocol

Platform partnerships: Cooperation agreements with social media platforms and news aggregators for instrumented deployment.

Ethics and privacy: IRB approval for minimal-risk observational study. No personally identifiable information collected. Opt-in telemetry with clear disclosure. Independent ethics board oversight.

Implementation: Browser extension and mobile app providing SCP verification. Telemetry logs: verification button taps, time spent, final decisions. Server logs: provenance queries, bundle downloads.

Rollout design: Clustered randomization by geographic region or user cohort to minimize spillover effects. Staged deployment over 6-month period.

Outcome measures:

- Verification engagement rates
- Time-to-verification decision
- Accuracy of verification decisions
- Sharing behavior for verified vs. unverified content
- Network effects on information propagation

C Implementation Artifacts

Upon journal acceptance, simulation code, experimental protocols, and anonymized analysis scripts will be released under open-source licenses with comprehensive documentation for reproduction studies.