The Shape of Trust: Designing Information Systems That Hold the Door Open

Preprint — Request for Feedback (General Online Edition)

Status: Public preprint, not yet peer-reviewed.

Purpose This edition is released for broad public feedback before formal venue-targeted submission. The goal is to stress-test claims, identify weak points, and improve clarity, falsifiability, and evidence quality.

How to read this version - Treat this as a working research artifact. - Distinguish between established claims, conjectures, and proposed tests. - Prioritize criticism that is specific and falsifiable.

What feedback is most useful 1) Which claims are underspecified or overclaimed? 2) Which predictions are not operationalized enough to test? 3) Where does the geometry-to-design mapping break or overreach? 4) Which references are missing from adjacent literatures? 5) Which parts are clear to non-specialists vs only to insiders?

How to respond Please reply publicly or directly with: - exact section number, - the claim you challenge, - why it fails (logic/evidence/operationalization), - and what observation would change your mind.

The Shape of Trust: Designing Information Systems That Hold the Door Open

Abstract

A digital system has direct access only to recorded events; everything else it carries is prediction. We define trust as the measurable gap between predicted and observed events that a viable system must hold to act ahead of confirmation. If this gap is forced to zero, systems stall. If it grows unbounded, systems field-fake: they continue producing coherent outputs while losing contact with reality. We formalize this gap in information-theoretic terms and propose a geometric instantiation on S^3 (unit quaternions), where trust corresponds to geodesic displacement sigma between predictive and observed state. We then derive design implications for institutions, businesses, and AI systems operating under mismatched clock rates between capital velocity, organizational mastery, and customer adaptation. We treat several stronger geometric correspondences as conjectures and provide falsifiable predictions rather than asserting closure. The practical discipline is simple: play games, derive religions. Run reality-settled processes first; codify only what survives repeated disconfirmation.

---

1. Introduction

Many contemporary systems emit confidence faster than they earn calibration. AI outputs are fluent before they are validated. Organizations revise strategy before they can sample customer effects. Institutions are pushed to settle questions on clocks that invalidate the sampling needed to settle them honestly.

This paper advances three claims.

1) Trust claim: trust is the carried gap between prediction and observation in event-driven systems.
2) Geometric claim: this gap can be computed as a state distance on a compact composition manifold; we use S^3 as a worked instantiation in a collaborative research line with da Silva (2026).
3) Design claim: robust systems preserve a bounded, continuously measured trust gap, maintain representational slack, and align operating speed with sampling capacity.

We do not claim finality. We provide a computational frame, explicit limits, and tests that could falsify parts of the synthesis.

---

2. Event-Log Ontology of Digital Systems

A digital system's state is the pair (event log, predictive model). It only knows what has been observed and recorded. Every non-observed state variable used for action is prediction.

Formally, let O_t be observed events up to time t and M_t be a predictive model over future events. Action requires selecting a_t before O_{t+1} is known. Therefore all viable systems act under unresolved prediction.

Two failure modes bound viability: - Zero-gap failure: requiring full settlement before action causes paralysis. - Unbounded-gap failure: letting unresolved prediction accumulate without measurement causes decoherence and eventual correction shock.

Trust, in this framework, is not sentiment. It is carried unresolved exposure between model-led action and posted events.

---

3. Trust as Measurable Surprise

Using Shannon surprise, information in an observed event e is I(e) = -log p_M(e), where p_M is model probability.

This yields a measurable trust account: - Predictions that later match observations amortize safely. - Predictions that mismatch accumulate measurable surprise debt.

Critical limit: surprise is model-relative. If an event class is absent from the model's support, the system cannot represent the disconfirmation cleanly. It aliases out-of-model events into in-model categories (often noise). The system then appears coherent internally while accumulating unmeasured external error.

Design consequence: learning systems must preserve representational slack, i.e., event-space capacity beyond currently fitted observations.

---

4. Geometric Instantiation on S^3

We instantiate state on S^3 (unit quaternions) for three practical properties: associative composition, order-sensitive composition, and bounded state magnitude. Under this instantiation, trust gap is geodesic distance sigma between predictive and observed state carriers.

What we claim here: - sigma is computable and bounded on S^3. - sigma provides a single operational scalar for carried prediction-observation displacement.

What we do not claim as settled: - that S^3 is the unique physically correct substrate for all trust-bearing systems, - that one threshold transfers unchanged across all social and technical domains.

We use a domain-calibrated critical-threshold hypothesis as a testable operating assumption, not as a universal constant. In current drafts we treat candidate threshold values as empirical priors to be estimated per domain rather than fixed a priori.

---

5. Exchange Geometry and Three Currencies (Conjectural)

Systems operate under three practical currencies: information, time, and value/energy.

We propose a conjectural mapping to Hamilton composition components: - scalar accumulation component -> information compounding, - linear transport component -> directional time progression, - cross-product component -> novel structure generation (candidate value creation term).

If this mapping survives testing, then exchange quality depends on carrier complementarity: parallel carriers produce near-zero cross term (transfer without creation), while complementary carriers produce larger cross terms (positive-sum creation).

Status: conjecture. We provide empirical predictions in Section 10.

---

6. Critical Regime: Holding the Door Open

A viable system must maintain trust gap inside a bounded active regime: - too low: no anticipatory action; - too high: unresolved exposure exceeds reconciliation capacity.

Operational discipline: - allow temporary complexity at transition moments, - force post-transition collapse to simpler settled state, - avoid chronic low-grade unresolved complexity.

This is the practical meaning of holding the door open: maintain enough unresolved space for adaptation without permitting runaway aliasing.

---

7. Coupled Loops and Irreversible Adaptation

Every viable system runs two coupled loops: - fast loop: perceptual integration of new events, - slow loop: consolidation of priors.

Define: - world-gap sigma_w = d(p_t, o_t), prediction vs new observation, - self-gap sigma_s = d(p_t, projection(m_t)), prediction vs model self-coherence.

Healthy learning requires both gaps to stay measurable and bounded under continued sampling. Field-faking occurs when internal self-coherence improves while world-grounding degrades.

Human counterpart: irreversible adaptation. People are shaped by repeated environments and are not fungible across contradictory role structures at zero cost. Systems that enforce interchangeability despite adaptation debt externalize harm and eventually lose capacity.

Design rule: if continuity requires sustained human contradiction, redesign the role topology before extracting more adaptation from people.

---

8. Capital Velocity vs Mastery Rate

Current economies run three desynchronized clocks: - capital allocation clock (near-instant), - organizational mastery clock (slower, sample-constrained), - customer adaptation clock (fragmented and bounded).

When capital clock outruns mastery clock, firms face a structural fork: - slow down and preserve calibration, - or preserve velocity by manufacturing unsampled customer compressions.

Most choose the second under prevailing incentives. This generates systemic near-fit interactions: products and messaging that are almost-right, repeatedly. Result: diffuse trust deficit, high churn, and market signals partially built from aliased priors.

This is not a moral accusation. It is a rate-mismatch diagnosis.

---

9. Design Principles

1) Keep event truth labels strict: observed vs predicted.
2) Measure trust gap continuously; do not backfill prediction as observation.
3) Preserve representational slack to keep disconfirmation observable.
4) Keep fast and slow loops coupled to world posting, not only internal verifier scores.
5) Match operating speed to sampling capacity; where impossible, shrink scope.
6) Protect talent sovereignty: people can initiate states without becoming system infrastructure.
7) Respect irreversible adaptation: dissolve roles before forcing chronic contradiction.
8) Remain losable: designers must permit disconfirmation inside their own systems.
9) Play games first; derive doctrine second.

---

10. Falsifiable Predictions and Evaluation Protocol

P1: Sampling-calibration link Organizations with higher measured customer sampling frequency should show lower prediction-observation gap in customer-model claims, and better retention persistence after subsidy removal.

P2: Pre-collapse trust-gap signal Major institutional failures should show measurable trust-gap growth before visible break events; event studies should detect increasing unresolved prediction load ahead of failure.

P3: Complementarity-value link Transactions between functionally complementary counterparties should produce more durable joint outcomes than transactions between near-parallel counterparties, controlling for sector and scale.

Minimal protocol: - define carrier representation per domain, - estimate sigma trajectories over time, - pre-register thresholds and validation windows, - publish misses and successful forecasts with equal weight.

---

11. Limitations

1) Substrate limitation: S^3 is an implementation choice with attractive properties, not proven universal substrate.
2) Threshold limitation: transferred critical thresholds require domain calibration; universal transfer is unproven.
3) Measurement limitation: organizational carrier definitions are under-specified and may be noisy.
4) Causal limitation: observed calibration improvements may be confounded by unobserved process quality.
5) Normative limitation: design principles can reduce harm but do not remove tradeoffs under hard resource constraints.

If central predictions fail, the framework should be decomposed and revised rather than defended as a block.

---

12. Conclusion

The central proposal is narrow and testable: trust is the carried, measurable gap between prediction and observation in systems that must act before full settlement. Good systems neither collapse this gap to zero nor let it drift unmeasured. They hold it in a bounded active regime, preserve disconfirmation pathways, and align speed with sampling reality.

The strongest geometric correspondences in this paper are offered as conjectures with failure conditions, not as doctrine. That distinction is the point. The discipline is recursive: run falsifiable games, then codify what survives.

Play games, derive religions.

References

Berezinskii, V. L. (1971). Destruction of long-range order in one-dimensional and two-dimensional systems possessing a continuous symmetry group. Soviet Physics JETP.

Kosterlitz, J. M., & Thouless, D. J. (1973). Ordering, metastability and phase transitions in two-dimensional systems. Journal of Physics C.

Da Silva, W. H. A. (2026). The Geometric Computer: Turing Completeness, Free Energy, and Learning in a Digital Brain on S^3. Open Research Institute.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nature Reviews Neuroscience, 11, 127-138.

Guinan, L. (2025). Conscious Systems: The Future of Human Coordination. Idea Nexus Ventures.

Hebbar, P., Manawat, Y., Verboomen, S., Ivanova, A., Palanimalai, S., Bhatia, K., & Baskaran, V. (2026). SIA: Self Improving AI with Harness & Weight Updates. arXiv:2605.27276.

Hurwitz, A. (1898). Über die Komposition der quadratischen Formen von beliebig vielen Variabeln. Nachrichten der Gesellschaft der Wissenschaften zu Göttingen.

Shannon, C. E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27.

Feedback Protocol (for this preprint round)

If you think this is wrong, excellent — show where it breaks. Best responses include: - a counterexample, - a better competing model, - or a test that would falsify this faster.

This preprint will be revised in public. Strong criticism improves the paper.