

The Tri-Organ Stack

A SEDP-Locked Architecture for Stable Facts, Stable Observers, and Stable Claims

Michael Zot

ZotBot Research Initiative

ORCID: 0009-0001-9194-938X

December 29, 2025

Abstract

Abstract. This note defines a minimal architecture for durable knowledge in a world of bounded observers. We SEDP-lock a Tri-Organ Stack composed of: (i) an *Objectivity Organ* \mathcal{O} that produces stable, redundantly readable records from quantum substrates, (ii) an *Active Inference Organ* \mathcal{A} that produces persistent bounded observers capable of acting and updating, and (iii) a *Witness Organ* \mathcal{W} that stabilizes claims under admissible adversarial challenges. We give operational definitions, a unified necessary-and-sufficient success predicate, lesion failures, minimality arguments, and witness tests. The result is a single system-level object that clarifies why replication can fail without new physics and why agreement requires more than observation, memory, or social consensus.

1 Why a stack is needed

Stable facts are not automatically trustworthy. Stable observers are not automatically correct. Agreement is not automatically robustness.

Durable knowledge requires three distinct sub-systems: a mechanism that produces public records, a mechanism that produces persistent agents, and a mechanism that converts records into adversarially stable claims. This note formalizes that triad as one composed architecture and links it to three companion papers that fully specify each organ [1, 2, 3].

2 SEDP in one paragraph

SEDP (System Emergence Discovery Protocol)

SEDP is a protocol for turning scattered mechanisms into a system-level object by forcing five locks. Definition Lock forces an operational definition. Success Predicate Lock forces a measurable criterion for success. Lesion Lock forces predicted failures when sub-functions are removed. Minimality Lock forces pruning to the smallest set that still works. Witness Lock forces diagnostics that decide success or failure in practice.

3 Phenomenon

Phenomenon

Operational durable knowledge is present when bounded agents can extract the same fact variable from disjoint records, persist long enough to repeat sampling and intervention, and converge on claims that remain stable under admissible adversarial challenges.

4 Definition Lock

Operational definition

The Tri-Organ Stack is present when independent bounded agents can (i) infer the same fact variable X from disjoint environment fragments, (ii) maintain themselves long enough to repeatedly sample and act, and (iii) converge on claims about X that remain stable under a defined class of admissible challenges, without requiring privileged access or coordination.

5 The three organs

5.1 Objectivity Organ \mathcal{O}

Role
\mathcal{O} succeeds when a variable X becomes operationally objective: independent observers can infer X from disjoint records without coordination and without materially disturbing the source. This requires (i) locally accessible redundancy, (ii) persistence under erasure and reset constraints, and (iii) scale stability of effective descriptions.

This organ formalizes the decoherence, redundancy, thermodynamic stabilization, and coarse-graining stack in a single necessary-and-sufficient condition [1]. The physics foundations include decoherence and einselection [4, 5], environment-as-witness objectivity [6], erasure bounds [7, 8], and scale stability via renormalization [9]. Known pitfalls in naive mutual-information witnesses and the role of SBS/Strong-QD structure are treated explicitly in the companion sources [10, 11, 12, 13].

5.2 Active Inference Organ \mathcal{A}

Role
\mathcal{A} succeeds when bounded observers persist over a horizon $[0, T]$ under uncertainty and control constraints, maintaining an internal state and policy that remain coherent and viable. Active inference provides a structured candidate mechanism via variational free energy minimization under a statistical boundary.

This organ is SEDP-locked in the companion paper [2], grounded in active inference and Free Energy Principle literature [14, 15, 16, 17]. The key point here is minimal: stable facts require stable observers capable of repeated sampling, intervention, and updating.

5.3 Witness Organ \mathcal{W}

Role
\mathcal{W} succeeds when candidate claims about X remain stable under an admissible class of adversarial challenges. It converts records into durable commitments by enforcing boundary rules, inference updates, challenge generation, confidence governance, and stabilization across time and turnover.

This organ is defined and SEDP-locked in the companion paper [3]. The general motivation aligns with falsifiability and adversarial testing norms [18], and with modern reproducibility and robustness concerns [19, 20].

6 Success Predicate Lock

Unified success predicate

Over horizon $[0, T]$, the Tri-Organ Stack exists if and only if there exist X , an environment partition, observer set, and an admissible challenge class such that:

$$\exists X, \{E_k\}, \{A_i\}, \mathcal{C} \text{ s.t. } \text{OBJ}(X) \wedge \text{AGT}(\{A_i\}) \wedge \text{WIT}(C, \mathcal{C}).$$

Here $\text{OBJ}(X)$ denotes \mathcal{O} success, AGT denotes \mathcal{A} success, and WIT denotes \mathcal{W} success for claims about X .

Operationally: facts must be extractable from disjoint records; observers must persist to repeat sampling and act; and claims must survive structured attempts to break them.

7 Lesion Lock

Organ-level lesion predictions

Lesion of \mathcal{O} : no stable public records, so convergence is impossible even with perfect observers.

Lesion of \mathcal{A} : no persistent observers, so records cannot be repeatedly sampled or used.

Lesion of \mathcal{W} : no adversarial filter, so claims drift with incentives and narratives rather than evidence.

Cross-cutting lesion signatures

Boundary failure yields confounds and injection.

Precision failure yields dogmatism or relativism.

Action failure yields correlation traps.

Stabilization failure yields drift and institutional amnesia.

8 Minimality Lock

Minimality argument

Two-organ subsets fail to produce durable knowledge: $\mathcal{O} + \mathcal{A}$ without \mathcal{W} yields unfiltered consensus vulnerable to systematic delusion. $\mathcal{O} + \mathcal{W}$ without \mathcal{A} yields procedures without persistent agents to run them. $\mathcal{A} + \mathcal{W}$ without \mathcal{O} yields adversarial debate without stable facts. Therefore, the triad is a minimal architecture for the stated phenomenon.

9 Witness Lock

Stack-level witnesses

Cross-fragment consensus: different agents infer the same X from disjoint records.
Intervention separation: actions split competing models predictably.
Adversarial robustness: red-team challenges fail to flip claims within admissible rules.
Time stability: claims persist across operator turnover and toolchain changes within tolerance.

10 Conclusion

The Tri-Organ Stack is a single SEDP-locked object that links stable facts, stable observers, and stable claims. It explains how replication can fail without new physics when convergence machinery is damaged. It also provides a diagnostic language for science, engineering, AI safety, and institutional epistemics by mapping failures to lesion signatures.

A Companion papers and canonical links

Companion papers

Objectivity Organ: GitHub PDF
Active Inference Organ: GitHub PDF
Witness Organ: GitHub PDF

References

- [1] M. Zot, *The Objectivity Organ: Definition, Criteria, and Failure Modes for Classical Objectivity*, ZotBot Research Initiative, GitHub PDF (2025). Link.
- [2] M. Zot, *The Active Inference Organ: A SEDP-Locked Account of Bounded Agency and Persistent Selves*, ZotBot Research Initiative, GitHub PDF (2025). Link.
- [3] M. Zot, *The Witness Organ: Definition, Criteria, and Failure Modes for Adversarially Stable Claims*, ZotBot Research Initiative, GitHub PDF (2025). Link.
- [4] W. H. Zurek, *Decoherence, einselection, and the quantum origins of the classical*, Rev. Mod. Phys. **75**, 715 (2003).
- [5] E. Joos et al., *Decoherence and the Appearance of a Classical World in Quantum Theory*, Springer (2003).
- [6] H. Ollivier, D. Poulin, and W. H. Zurek, *Objective Properties from Subjective Quantum States: Environment as a Witness*, Phys. Rev. Lett. **93**, 220401 (2004).
- [7] R. Landauer, *Irreversibility and Heat Generation in the Computing Process*, IBM J. Res. Dev. **5**, 183 (1961).
- [8] A. Bérut et al., *Experimental verification of Landauer's principle linking information and thermodynamics*, Nature **483**, 187 (2012).
- [9] K. G. Wilson, *Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture*, Phys. Rev. B **4**, 3174 (1971).
- [10] J. K. Korbicz, *Roads to objectivity: Quantum Darwinism, Spectrum Broadcast Structures, and Strong Quantum Darwinism*, Quantum **5**, 571 (2021).
- [11] T. P. Le and A. Olaya-Castro, *Strong Quantum Darwinism and Strong Independence are Equivalent to Spectrum Broadcast Structure*, Phys. Rev. Lett. **122**, 010403 (2019).
- [12] D. A. Chisholm et al., *Witnessing Objectivity on a Quantum Computer*, Quantum Sci. Technol. **7**, 015022 (2022).
- [13] D. A. Chisholm, L. Innocenti, and G. M. Palma, *Importance of using the averaged mutual information in quantum Darwinism: definitions and examples*, Phys. Rev. A **110**, 012218 (2024).
- [14] K. Friston, *The free-energy principle: a rough guide to the brain*, Trends Cogn. Sci. **13**(7), 293–301 (2009).
- [15] K. Friston, *The free-energy principle: a unified brain theory?*, Nat. Rev. Neurosci. **11**, 127–138 (2010).
- [16] T. Parr and K. Friston, *Generalised free energy and active inference*, Biol. Cybern. **113**, 495–513 (2019).
- [17] C. L. Buckley et al., *The free energy principle for action and perception: a mathematical review*, J. Math. Psychol. **81**, 55–79 (2017).
- [18] K. Popper, *The Logic of Scientific Discovery*, Hutchinson (1959).
- [19] J. P. A. Ioannidis, *Why Most Published Research Findings Are False*, PLoS Med. **2**(8), e124 (2005).
- [20] Open Science Collaboration, *Estimating the reproducibility of psychological science*, Science **349**(6251), aac4716 (2015).