

Round 26

What Can the Watchers Ever Know?

Minimum Observables for Detecting Omitted Change

A formal identifiability analysis and observation-sufficiency protocol

17 October 2026

Research paper / OSC-0.1-development / E26-A finite-model diagnostic

One exact finite-model analysis; no production, language-model, or answer-quality benchmark.

Contents

This publication separates frozen observations, current-state claims, external findings, and Enceavra hypotheses.

Result boundary	5
Abstract	5
1. Archive continuity and corrected round number	6
2. Research question, novelty, scope, and exclusions	7
2.1 Primary question	7
2.2 Novelty against the archive	7
2.3 In scope	7
2.4 Exclusions	7
3. Definitions and evidence semantics	8
4. Literature map and competing positions	8
4.1 Impossibility results are conditional, not pessimistic slogans	8
4.2 Knowledge is defined by possible worlds that look the same	9
4.3 Diagnosability asks whether faults leave distinguishable observations	9
4.4 Failure detectors add information with explicit accuracy and completeness properties	9
4.5 Accountability and transparency cover observed or committed actions	9
4.6 More agent discussion can help or harm	9
4.7 Active sensing offers a limited, testable analogy	10
5. Claim and evidence map	10
6. Formal model	10
6.1 Worlds, targets, and observations	10
6.2 Separation matrix	11
6.3 Decision semantics	11
7. Propositions and proofs	11
Proposition 1: observational indistinguishability limit	11
Proposition 2: root-disjoint lower bound	11
Proposition 3: health-only non-identification	12
Proposition 4: three exhaustive ways to repair an ambiguous claim	12
8. Diagnostic E26-A	12
8.1 Purpose	12
8.2 Root model	12
8.3 Observable semantics	12
8.4 Enumeration	13
8.5 Named package results	13

8.6 Minimum identifying sets	14
8.7 Root-lineage ablation	14
8.8 Minimal blinding cuts	14
8.9 Verification and reproducibility	14
8.10 Evidence boundary	14
9. Minimum additional requirement by omission class	15
10. Observation Sufficiency Certificate	15
10.1 Purpose	15
10.2 Proposed schema	16
10.3 Protocol	16
10.4 What the certificate must not do	16
11. Worked example	17
12. Brain-inspired hypothesis, kept narrow	17
13. Alternative explanations and rival designs	18
13.1 Conservative abstention may explain apparent safety	18
13.2 An external channel may help because it is better, not because it is independent	18
13.3 Root metadata may be wrong	18
13.4 Active probes may perturb the system	18
13.5 The finite model may omit the actual failure	18
13.6 A probabilistic policy may be more useful than a zero-error certificate	18
13.7 End-to-end revalidation may dominate formal selection	18
14. Falsifiable hypotheses	19
15. Proposed experiment E26-B	19
15.1 Prerequisites	19
15.2 Strata	19
15.3 Policies	20
15.4 Controlled total compute and observation budget	20
15.5 Baselines and ablations	20
15.6 Primary metrics	21
15.7 Regression and harmful-change metrics	21
15.8 Analysis plan	21
16. Expected failure modes of OSC	21
17. Limitations	22
18. Decision and next-round handoff	22
18.1 Round 26 decision	22
18.2 Preferred Round 27	22

18.3 Fallback if access remains unavailable 23

19. Reproducibility record 23

20. Source inspection register 23

21. Full sources 24

22. Closing statement 25

Result boundary

Round 26 asks when a collaborative AI system can know that a quiet evidence window really means no relevant change. It proves a general, model-independent limit: if an admissible changed world and an admissible unchanged world produce exactly the same observations, no decision rule over those observations can be correct in both worlds. The only remedies are to add an observable that separates the worlds, exclude one world with an explicit trust assumption, or weaken the claim.

The round then executes E26-A, a deterministic finite-state analysis over seven possible failure roots and eight observable types. The analysis enumerates every fault set up to budgets of zero, one, two, and three failed roots, tests every observable subset, lists all opposite-target worlds that remain observationally identical, and finds inclusion-minimal and minimum-cost separating sets. The declared state space contains 2, 16, 58, and 128 worlds at those budgets, respectively. Eight executable tests passed.

Within this declared model, one target-sensitive channel is sufficient with no faults, two root-disjoint target-sensitive channels are necessary and sufficient against one arbitrary silent root failure, three are necessary and sufficient against two, and four are necessary and sufficient against three when the fourth witness is genuinely independent. When the nominal witness is allowed to share the issuer's root, no available set identifies change against three failed roots. This is an exact result for the finite model and a direct instance of the root-disjointness theorem in Section 7. It is not an estimate of real incident frequency or AI performance.

The Round 25 SACP observable set still has eight opposite-target ambiguity pairs under the one-fault universe. Adding one independently rooted external commitment reduces that count to zero under the same universe. Under two faults, SACP plus one external commitment leaves 26 ambiguity pairs; adding an independently rooted behavior probe reduces the count to zero. These counts weight every enumerated fault set equally and must not be read as prevalence-weighted rates.

No production incident was sampled. No live repository webhook, deployment control plane, database, provider account, runtime, user traffic, or customer data was queried or changed. No language model was benchmarked. No new Enceavra answer benchmark was run. The proposed certificate was not installed. No release gate, claim state, retention policy, or production behavior changed.

Abstract

Round 25 showed that liveness challenges, sequence commitments, and self-monitoring can detect many ways in which missing change evidence masquerades as stability. It also recorded five designed misses: selective non-emission without an external commitment and verifier common-mode failure without an independent challenge path. Its prescribed fallback, when a real multi-authority shadow trial with hidden adjudication is unavailable, was a formal indistinguishability analysis rather than a larger authored fixture suite.

Round 26 takes that fallback. Its research question is: **For each omitted-change failure class, which unchanged and changed worlds remain observationally identical under a given observer set, and what is the minimum additional independently rooted observable or trust assumption needed to separate them?**

The central object is an observation function. A world contains the target fact, such as whether a relevant authority change occurred, plus latent fault state. A selected observer set maps that world to an observable vector. Two worlds are indistinguishable when the vectors are equal. The target is identifiable only when every observational equivalence class contains one target value. This makes the boundary operational: more messages, agents, votes, or checks help only if they split an ambiguity class that crosses the target boundary.

Primary distributed-systems results establish why guarantees depend on timing, authentication, failure, and knowledge assumptions [S05-S10]. Diagnosability research asks whether faults can be inferred from

observable event sequences and extends the idea to active diagnostic actions [S09, S11]. PeerReview, Certificate Transparency, and TUF show how signed records, witnesses, append-only commitments, versions, and expiry can expose some faults while retaining explicit trust limits [S12-S14]. Mixed multi-agent language-model evidence shows both benchmark gains from debate and systematic cases in which discussion propagates errors or verification fails [S15-S17]. Together these sources support the need to reason about information structure before spending more computation on collaboration. They do not establish the proposed Enceavra mechanism.

Round 26 proposes the **Observation Sufficiency Certificate (OSC)**. Before a system certifies `NO_RELEVANT_CHANGE`, the OSC names the target predicate, threat model, failure roots, selected observables, root-lineage assumptions, separated world pairs, residual indistinguishable pairs, verifier trusted computing base, cost, and expiry. If any residual equivalence class contains both changed and unchanged worlds, the only admissible result is `UNKNOWN_WITH_RESIDUAL_AMBIGUITY`, not stability.

The useful brain-inspired idea is limited to active sensing. Biological active sensing is described as a closed loop in which an observer estimates task-relevant state and a planner chooses actions expected to obtain informative observations [S18]. Enceavra's testable analogue is to choose the next probe by expected ambiguity reduction per unit cost. This is not a claim that the protocol copies a brain. A fixed-budget experiment can reject the analogy if adaptive probes do not reduce residual ambiguity or harmful false-stability decisions relative to fixed probes.

1. Archive continuity and corrected round number

The automation brief retained the historical label Round 11. Before topic selection, the live Lovable project was inspected at commit `f1197fa663970737c3da35bc8e98c63c60f59c78`. The Research overview, rounds index, registry, full Round 25 and Round 24 manuscripts, and the relevant Round 10 and Round 9 routes show completed work through Round 25. The correct next sequential number is therefore **Round 26** [S01-S04].

The carried program state is:

Round	Contribution carried forward	Boundary retained in Round 26
9	Reusable procedures should be learned only from verified experience.	E26-A is a formal diagnostic, not training data or a learned procedure.
10	Agents should exchange recoverable evidence, revise selectively, compare against strong same-cost baselines, and treat brain-inspired ideas as hypotheses.	Round 26 studies what can be known before revision and proposes a fixed-cost falsification plan.
24	Evidence must preserve authority lineage; two analyses of one root are not independent authorities.	Root labels remain assumptions until their dependency graph is verified.
25	Empty windows require liveness, checkpoint, transport, and reconciliation evidence; SACP detected most authored gaps but missed selective non-emission and verifier common mode.	Round 26 formalizes the residual blind spots instead of expanding the authored trace suite.

Round 25 preferred a preregistered E25-B shadow window over real multi-authority observations with hidden independent adjudication. That access was not available in this run. Round 25 explicitly prescribed the present fallback: enumerate omission classes that remain undetectable under each observer set and prove the minimum new trust assumption needed to separate them [S02].

2. Research question, novelty, scope, and exclusions

2.1 Primary question

For each omitted-change failure class, which changed and unchanged worlds produce the same observable trace under the selected observer set, and what is the minimum new observable, independence condition, timing assumption, or claim restriction needed to distinguish them?

The decision target is deliberately narrow:

$T(w) = 1$ if a relevant authority change occurred in the declared window and claim scope; otherwise $T(w) = 0$.

Round 26 does not try to identify every root cause. It first asks the simpler and logically prior question: can the target fact be identified at all from the evidence available to the system?

2.2 Novelty against the archive

Round 24 asks where received evidence originated. Round 25 asks whether configured observation paths were live, fresh, continuous, and reconciled. Round 26 asks whether those observations contain enough information to distinguish the target worlds even if every implemented check behaves exactly as specified.

This is not another evidence-exchange protocol, selective-revision policy, or procedural-learning method. The new object is the equivalence class induced by an observer set. It provides a stopping rule for verification effort: a check has zero value against a candidate failure when it produces the same output in both members of the unresolved world pair.

Three adjacent directions were considered and deferred:

Candidate	Reason deferred
Confidence calibration	Important for the Round 10 roadmap, but a calibrated confidence value cannot recover a fact absent from every observed channel. Identifiability is prior.
Fixed-budget verification allocation	Round 26 supplies the missing objective: spend verification only on actions that split live ambiguity classes. A full learned allocator belongs after this objective is validated.
Provenance-preserving conflict memory	Rounds 22-24 already address claim state and authority lineage after evidence exists. The unresolved issue is missing information, not storage of observed conflict.

2.3 In scope

- Binary target identifiability for a declared authority, claim, and time window.
- Silent issuer omission, transport loss, observer failure, stale or false-clean checkpoints, external-channel failure, probe blind spots, common-origin witnesses, and verifier failure.
- Exact observational equivalence, minimal separating sets, root-disjointness, and claim narrowing.
- A finite root-fault model with transparent assumptions and executable enumeration.
- A proposed OSC data structure and three-way decision rule.
- A prospective fixed-budget experiment with natural and sandbox strata kept separate.

2.4 Exclusions

- Estimating the prevalence or probability of any fault.
- Proving the correctness of an authority, observer, cryptographic primitive, clock, network, or verifier.
- Continuous-time behavior, arbitrary Byzantine strategies, or every possible interleaving.
- Product answer accuracy, calibration, safety, latency, user trust, or economic value.

- Deploying a monitor or changing a claim's production status.
- Literal brain copying.

3. Definitions and evidence semantics

Term	Round 26 definition	Common invalid substitution
World	A target value plus all latent states relevant to observation.	A single log line or agent response.
Observable	A value available to the decision rule, including a signed record, explicit clean close, missing status, or challenge result.	A component that exists but whose output is not independently available.
Observer set	The named observables and their declared root dependencies.	A count of agent or service names.
Observation function	The mapping from a world to the selected observable vector.	A confidence score about the world.
Indistinguishable worlds	Two worlds with identical selected observable vectors.	Worlds that merely look similar to a human reviewer.
Identifiable target	The target is constant within every observational equivalence class.	High average accuracy on sampled traces.
Failure root	A dependency whose failure can blind one or more observables together.	A process name, model name, or network address.
Root-disjoint channel	A target-sensitive channel whose relevant failure dependencies do not overlap another channel under the threat model.	A duplicate agent reading the same source.
Health observable	A signal about path status, such as heartbeat, acknowledgment, or canary.	Direct evidence that the target did or did not change.
Target-sensitive observable	A signal that differs between at least one changed and unchanged world.	Any additional message.
Residual ambiguity	An equivalence class containing both target values after all selected observations.	A low-confidence but still certifiable no-change result.
Trust assumption	A world excluded from the admissible set, such as at most one failed root or an honest external signer.	An unrecorded belief that a service is probably independent.

Three levels of claim must remain separate:

- 1 **NO_RECEIVED_CHANGE**: the selected receipt stream contains no relevant change.
- 2 **NO_RECORDED_CHANGE**: a named authority commitment explicitly closes the window as clean.
- 3 **NO_ACTUAL_RELEVANT_CHANGE**: the target authority did not undergo a covered change.

The first can be true when transport fails. The second can be true when the issuer selectively omits. The third requires assumptions or independent observations that rule out both. OSC prevents a system from silently upgrading one level into another.

4. Literature map and competing positions

4.1 Impossibility results are conditional, not pessimistic slogans

Lamport, Shostak, and Pease show that Byzantine agreement depends on the number and behavior of faulty participants and on whether messages are authenticated [S05]. Fischer, Lynch, and Paterson show a nontermination possibility for deterministic consensus in a fully asynchronous model with one crash fault [S06]. Dwork, Lynch, and Stockmeyer recover solvability under declared partial-synchrony assumptions [S07].

The shared lesson is not that reliable collaboration is impossible. It is that a guarantee is meaningful only relative to a model. A Round 26 certificate must therefore name timing, fault, authentication, and independence assumptions rather than bury them in implementation prose.

4.2 Knowledge is defined by possible worlds that look the same

Halpern and Moses formalize knowledge in distributed systems using points that an agent or group cannot distinguish. They show that strong common knowledge is unattainable in many practical systems with temporal imprecision, while weaker variants can support useful coordination [S08].

Round 26 uses the indistinguishability construction, not the full epistemic logic. Its practical consequence is similar: when both a changed and unchanged world remain possible under the same view, the system does not know which target value holds. The appropriate response may be a weaker statement, such as `no recorded change through checkpoint q`, rather than an unsupported global stability claim.

4.3 Diagnosability asks whether faults leave distinguishable observations

Sampath and colleagues define diagnosability for discrete-event systems in terms of whether fault occurrences can be inferred from observable behavior within a bound [S09]. Their later active-diagnosis work studies control actions that make otherwise ambiguous faults distinguishable [S11].

This is the closest formal precedent for Round 26. The mapping is not exact: Enceavra's target is a claim-relevant change, its observers can include AI agents and external services, and the finite model is not a general automaton over unbounded traces. Still, diagnosability supplies the right question: what information is present in the projection that the monitor can see?

4.4 Failure detectors add information with explicit accuracy and completeness properties

Chandra and Toueg model failure detectors as potentially incorrect information about failures and classify them by completeness and accuracy [S10]. The detector is an oracle abstraction with stated properties, not proof that a timeout reveals the underlying cause.

Round 26 similarly treats heartbeats, acknowledgments, and canaries as health information. They can turn false certainty into `UNKNOWN`, which is valuable. They do not by themselves make a hidden semantic change identifiable.

4.5 Accountability and transparency cover observed or committed actions

PeerReview maintains signed records and uses correct witnesses to expose deviations that are observed by a correct node [S12]. RFC 9162 uses append-only logs, signed timestamps, inclusion proofs, consistency proofs, and monitoring to make recorded certificate activity auditable [S13]. TUF uses signed roles, monotonic versions, expiry, and threshold rules to expose rollback and freeze under a declared threat model [S14].

These mechanisms support checkpoints, root separation, freshness, and post hoc accountability. Their stated limits matter just as much. A correct log of incomplete input is still incomplete. A monitor must exist and observe the commitment. A set of witnesses is only as independent as its actual failure roots.

4.6 More agent discussion can help or harm

Du and colleagues report benchmark improvements from multi-agent debate across selected reasoning and factuality tasks [S15]. Wynn, Satija, and Hadfield report settings in which debate decreases accuracy and causes stronger agents to revise correct answers toward incorrect peers [S16]. Cemri and colleagues identify multi-agent failures including information withholding and incomplete or incorrect verification in the systems they studied [S17].

The evidence is mixed, not contradictory. Interaction can expose useful information when agents possess or elicit different evidence. It can amplify error when agents share a blind source, accept persuasive mistakes,

or fail to verify. Round 26 predicts that agent count alone should not improve an omission guarantee when all agents inhabit the same observational equivalence class.

4.7 Active sensing offers a limited, testable analogy

Yang, Wolpert, and Lengyel describe active sensing as a closed loop between an observer that estimates task-relevant state and a planner that chooses actions expected to obtain informative sensory input [S18].

Round 26 borrows one design hypothesis: choose probes based on their expected ability to split the current ambiguity class, not on a fixed schedule alone. The analogy is rejected if this rule fails under matched observation cost. No claim is made that an AI assurance system implements neural computation.

5. Claim and evidence map

Claim	Evidence	Status	Boundary or falsifier
If opposite-target worlds have identical selected observations, no decision rule over those observations can be correct in both.	Proposition 1, a direct function argument; epistemic precedent [S08].	Established mathematically.	A purported rule that returns different values for identical inputs is not a function over those inputs.
Under the stated silent-root model, tolerating up to f arbitrary failed roots requires at least $f + 1$ root-disjoint target-sensitive channels.	Proposition 2 and exhaustive E26-A enumeration.	Established for the stated model class.	A counterexample satisfying all stated assumptions with f or fewer roots.
Health observables can justify abstention without identifying the target.	Proposition 3; Chandra-Toueg distinction [S10]; E26-A.	Established for health signals conditionally independent of the target.	A health signal whose semantics are actually target-sensitive falls outside the premise.
SACP alone is structurally ambiguous under one silent issuer failure in E26-A.	Eight opposite-target pairs and one false-clean changed world in the one-fault enumeration.	Exact finite-model result.	A different SACP semantics or threat model can change the result.
SACP plus one independent external commitment identifies the target under one failed root in E26-A.	Zero ambiguity pairs across 16 worlds.	Exact finite-model result.	Shared roots, more than one fault, false root metadata, or model mismatch.
A witness label without root independence is not protective.	Independent variant has zero one-fault ambiguity pairs for SACP plus witness; correlated variant has five.	Exact finite-model result and design warning.	Verified independent lineage could satisfy the independent variant.
OSC will reduce harmful false-stability certifications in real multi-authority windows at fixed budget.	Mechanistic rationale only.	Enceavra hypothesis, untested.	E26-B shows no improvement or worse regression-adjusted utility.
Adaptive ambiguity-splitting probes outperform fixed probes at equal cost.	Active-diagnosis and active-sensing precedents [S11, S18].	Enceavra hypothesis, untested.	Fixed or random probing matches or outperforms on preregistered metrics.

6. Formal model

6.1 Worlds, targets, and observations

Let W be the admissible world set. A world includes the target fact, latent authority state, and fault state. Let $T: W \rightarrow \{0, 1\}$ be the target predicate. Let S be a selected set of observables. The observation function is:

$$O_S: W \rightarrow Y_S$$

where Y_S is the Cartesian product of each selected observable's categorical outputs.

Define an equivalence relation:

$w \sim_S w'$ if and only if $0_S(w) = 0_S(w')$.

The target is identifiable under S when every equivalence class is target-pure:

For all w, w' in W , if $w \sim_S w'$, then $T(w) = T(w')$.

6.2 Separation matrix

For every admissible pair $p = (w_0, w_1)$ with $T(w_0) \neq T(w_1)$, define:

$M[p, s] = 1$ when observable s produces different values in the two worlds, and 0 otherwise.

An observer set identifies the target exactly when it hits every opposite-target pair:

For every p , at least one selected s has $M[p, s] = 1$.

With sensor costs c_s , choosing a least-cost identifying set is a weighted hitting-set problem. Round 26 solves the tiny instance by exhaustive enumeration so that no optimizer assumptions obscure the result.

6.3 Decision semantics

OSC uses three outputs:

- **CHANGE**: at least one valid selected channel reports a covered change.
- **NO_RELEVANT_CHANGE**: all target-crossing pairs are separated under the declared model and the observed vector lies only in target-zero worlds.
- **UNKNOWN_WITH_RESIDUAL_AMBIGUITY**: the current equivalence class contains both target values, a required observation is missing, or a certificate assumption is not met.

Abstention is not evidence of a change. It preserves the distinction between epistemic uncertainty and contradiction.

7. Propositions and proofs

Proposition 1: observational indistinguishability limit

Let w_0 and w_1 be admissible worlds such that $T(w_0) \neq T(w_1)$ and $0_S(w_0) = 0_S(w_1)$. No deterministic decision function d over 0_S can return the correct target value in both worlds.

Proof. Because the observable vectors are equal, the function receives the same input in both worlds. A function returns the same output for the same input. The target values differ, so that single output is wrong in at least one world. QED.

A randomized rule does not escape the guarantee boundary. It induces the same output distribution for the same observable vector, so it cannot assign probability one to different correct values in the two worlds. Randomization may trade error between worlds; it does not identify the hidden fact.

Proposition 2: root-disjoint lower bound

Assume each target-sensitive channel has one declared failure root. A failed root may force every channel under that root to emit exactly its quiet-world value. Under up to f arbitrary failed roots, zero-error identification of a change requires at least $f + 1$ target-sensitive root-disjoint channels.

Necessity. With at most f roots, an adversary can fail every available root and make all selected channels emit quiet values. The changed world is then indistinguishable from a quiet world. Proposition 1 applies.

Sufficiency under the premise. With $f + 1$ root-disjoint channels and at most f failed roots, at least one channel survives. If every surviving channel truthfully reports the target, a changed world produces at least one change observation, while a quiet world does not. QED.

The sufficiency statement is intentionally conditional. If a surviving channel lacks coverage, shares an undeclared dependency, or has an incorrect verifier, it is not a surviving truthful root under the premise.

Proposition 3: health-only non-identification

Suppose health observable `h` has the same value in a changed and unchanged world whenever their fault state is the same. Adding only `h` cannot split that same-fault opposite-target pair.

Proof. The pair has equal prior observations and equal `h`. Their extended vectors remain equal. QED.

Heartbeats, queue acknowledgments, and canaries remain useful. They can prevent a missing path from being interpreted as a clean close, turning `NO_RELEVANT_CHANGE` into `UNKNOWN`. They cannot show that a semantically hidden change occurred when their health state is normal.

Proposition 4: three exhaustive ways to repair an ambiguous claim

For an ambiguous opposite-target pair, a sound system must do at least one of the following:

- 1 add an observable with different outputs in the pair;
- 2 add a justified assumption that removes one member from the admissible world set;
- 3 weaken the target so both worlds have the same target value.

These are exhaustive because identifiability depends only on the world set, target function, and observation function. Changing none of them leaves the pair unchanged.

8. Diagnostic E26-A

8.1 Purpose

E26-A checks the internal consequences of the formal definitions and exposes the smallest failure cuts for named observable packages. It is a finite design-space analysis, not a production benchmark and not an authored sequence-fixture expansion.

8.2 Root model

Seven binary failure roots are enumerated:

Root	Worst-case effect in the model
<code>issuer</code>	The primary issuer silently omits the change and produces a false-clean checkpoint.
<code>transport</code>	Receipt and checkpoint transport fail; the acknowledgment is missing.
<code>observer</code>	Receipt and checkpoint observation fail; the heartbeat is missing.
<code>verifier</code>	Primary receipt and checkpoint evaluation are blinded; the independent canary reports failure.
<code>external</code>	The external commitment silently returns its quiet value.
<code>probe</code>	The behavior probe silently returns its quiet value because coverage or execution is blind.
<code>witness</code>	The witness silently returns its quiet value. In the correlated variant, issuer failure also blinds it.

This is a root model, not a component count. Receipt and checkpoint share several primary roots. External commitment, probe, and witness are independent only in the declared independent variant.

8.3 Observable semantics

Eight categorical observables are implemented:

Observable	Outputs	Target-sensitive	Main role
receipt	change, none	Yes	Received event evidence.
checkpoint	change, clean, missing	Yes	Explicit window close and primary commitment.
heartbeat	ok, missing	No	Observer liveness.
transport_ack	ok, missing	No	Transport health.
verifier_canary	ok, failed	No	Primary verifier health.
external_commitment	change, clean	Yes	Separately rooted authority commitment.
behavior_probe	change, clean	Yes	End-to-end active observation.
independent_witnesses	change, clean	Yes	Separately rooted witness, or a correlated mirror in the ablation.

Silent failure is intentionally adversarial: it produces the same value as a healthy quiet world. Transport, observer, and verifier health signals can cause **UNKNOWN**, but a silent issuer omission keeps health green.

8.4 Enumeration

For each maximum fault budget f , the script enumerates every subset of the seven roots with cardinality at most f , then crosses each set with both target values. The resulting world counts are:

Maximum failed roots	Fault sets	Worlds
0	1	2
1	8	16
2	29	58
3	64	128

For each of the 255 nonempty observable subsets, the analysis builds the exact observation vector for every world, lists every changed-quiet pair with equal vectors, finds inclusion-minimal identifying sets, and calculates minimum sets under a transparent illustrative cost model. Primary sensors cost one unit; separately rooted external, probe, and witness channels cost three units each.

The cost units are not dollars, tokens, or latency. They are a sensitivity device. The structural zero-versus-nonzero ambiguity result does not depend on them.

8.5 Named package results

The following table uses the independent-witness variant. Counts are exact for the enumerated worlds and unweighted by real-world probability.

Package and fault budget	Changed worlds detected	Changed worlds unknown	Changed worlds falsely clean	Opposite-target ambiguity pairs	Structurally identifying
Receipt absence baseline, $f \leq 1$	4 of 8	0	4	32	No
Round 25 SACP, $f \leq 1$	4 of 8	3	1	8	No
SACP plus external commitment, $f \leq 1$	8 of 8	0	0	0	Yes
SACP plus external, $f \leq 2$	25 of 29	3	1	26	No
SACP plus external plus probe, $f \leq 2$	29 of 29	0	0	0	Yes
SACP plus external plus probe, $f \leq 3$	60 of 64	3	1	48	No

SACP improves on receipt absence by converting transport, observer, and verifier failures into **UNKNOWN**. Its remaining one-fault false-clean case is issuer omission, exactly the residual class identified by Round 25. One external root closes that class under the one-fault assumption. A second external root is needed under two arbitrary failures.

8.6 Minimum identifying sets

Maximum failed roots	Independent witness: minimum target roots	Minimum illustrative cost	Example minimum-cost set	Correlated witness result
0	1	1	<code>checkpoint</code>	Same.
1	2	4	<code>checkpoint + external_commitment</code>	Still possible with main plus external or probe.
2	3	7	<code>checkpoint + external_commitment + behavior_probe</code>	Still possible with the three genuinely distinct roots.
3	4	10	<code>checkpoint + external + probe + witness</code>	Impossible with the available sensors because the witness collapses into the issuer root.

The exhaustive output contains every inclusion-minimal set, not only the examples in this table.

8.7 Root-lineage ablation

With SACP plus a witness and a one-fault budget:

Witness model	Ambiguity pairs	Changed false-clean worlds	Identifying
Genuinely independent witness	0	0	Yes
Witness shares issuer root	5	1	No

The code and message count are unchanged. Only the dependency graph changes. This is the clearest demonstration that nominal diversity is not epistemic diversity.

8.8 Minimal blinding cuts

- SACP has size-one structural cuts at issuer, transport, observer, and verifier. The latter three may produce **UNKNOWN** rather than false-clean when their health signals are included.
- SACP plus external commitment has size-two cuts pairing the external root with issuer, transport, observer, or verifier.
- SACP plus external commitment and behavior probe has analogous size-three cuts.

The distinction between a structural cut and a harmful false-clean decision matters. A visible health failure can leave the target unidentified while still preventing an unsafe certificate.

8.9 Verification and reproducibility

Eight deterministic unit tests passed. They cover the Round 25 issuer-omission pair, one- and two-root lower bounds, witness correlation, health-only insufficiency, named package guarantees, and output generation. The bundle includes the executable model, tests, complete ambiguity pairs, minimal sets, blinding cuts, package metrics, source inspection ledger, and SHA-256 manifests.

8.10 Evidence boundary

E26-A establishes logical consequences of its own semantics. It does not establish that seven roots are exhaustive, that each declared root is independent in a deployed system, that every failure produces the modeled output, or that any fault set is likely. Its exactness is internal, not ecological.

9. Minimum additional requirement by omission class

Failure class	Why current view is ambiguous	Minimum separating move	Residual assumption
Selective non-emission	A changed issuer and an unchanged issuer both produce clean primary observations.	Add one target-sensitive channel outside the issuer root, or weaken the claim to no issuer-recorded change .	The added channel covers the target and does not share the issuer failure.
Transport drop or indefinite delay	No receipt can mean no event, dropped event, or delayed event.	Add an authenticated acknowledgment or independent path plus a delivery-time assumption; otherwise return UNKNOWN .	A bound or eventual-delivery property is trusted [S07, S10].
Observer crash	Missing evidence can mean quiet or dead observer.	Add liveness or a second observer to prevent false-clean. To identify the target, add a target-sensitive independent channel.	Liveness signal itself is observed and timely.
Freeze or rollback	Old signed state can look current.	Add monotonic version, expiry, persistent prior state, and consistency check [S13-S14].	Clock, key, and retained-state assumptions hold.
Equivocation or split view	Each consumer sees an internally valid but different history.	Add gossip or witnesses that compare commitments across views.	At least one comparison path does not share the equivocating root.
Common-origin mirrors	Multiple observers repeat the same omission.	Verify root lineage and add a genuinely separate authority, probe, or witness.	Independence metadata matches operational dependencies.
Verifier common mode	Evidence exists but the only verifier suppresses or misreads it.	Replay immutable raw evidence through an independently implemented verifier, or narrow the trusted computing base.	Raw evidence survives outside the primary verifier and the second verifier is not correlated.
Probe blind spot	A probe exercises a path unaffected by the hidden change.	Bind probes to explicit coverage obligations and add a different target-sensitive route.	Coverage mapping is correct.
Joint failure of every target root	Every selected channel emits its quiet value.	Add a new trust root or weaken the claim.	No receipt-only protocol can solve this class.

10. Observation Sufficiency Certificate

10.1 Purpose

The OSC is a proof obligation attached to a negative change claim. It does not prove the world is safe. It records why the selected observations are sufficient under a specific threat model and exposes what remains unknown.

10.2 Proposed schema

```

osc_version: 0.1-development
certificate_id: osc:<claim>:<authority>:<window>
target:
  predicate: relevant_change_occurred
  claim_id: C21-CUR-TERMINAL
  authority: provider_contract
  window: [start, end]
  relevance_rule_digest: sha256:...
threat_model:
  maximum_failed_roots: 1
  timing: checkpoint_due_within_15m
  excluded_worlds:
    - reason
roots:
  primary:
    dependencies: [issuer, transport, observer, verifier]
  external_policy_log:
    dependencies: [external_signer, external_transport, verifier_2]
observables:
  - id: primary_checkpoint
    root: primary
    outcome: clean
    freshness: pass
  - id: external_commitment
    root: external_policy_log
    outcome: clean
    freshness: pass
separation:
  candidate_pair_count: 8
  unresolved_pair_count: 0
  matrix_digest: sha256:...
verifier_tcb:
  evaluator: osc-checker-v0.1
  raw_evidence_digest: sha256:...
decision: NO_RELEVANT_CHANGE
expires_at: timestamp
limitations:
  - no guarantee beyond one failed root
  - no guarantee outside the declared relevance rule

```

10.3 Protocol

- 1 Define the target predicate, authority, relevance rule, and time window before observing the result.
- 2 Enumerate or reference the admissible opposite-target world pairs under the threat model.
- 3 Register each observable with outputs, coverage, failure dependencies, freshness, and root lineage.
- 4 Build the separation matrix and select a cost-bounded set that hits every target-crossing pair.
- 5 Collect signed or reproducibly addressable raw observations.
- 6 Evaluate health, freshness, root-lineage, and target-sensitive outcomes without silently treating `missing` as `clean`.
- 7 Independently verify the certificate when verifier common mode is in scope.
- 8 Emit `NO_RELEVANT_CHANGE` only if the realized observation class is target-pure. Otherwise emit `CHANGE` or `UNKNOWN_WITH_RESIDUAL_AMBIGUITY`.
- 9 Expire the certificate when the window, threat model, root map, relevance rule, or verifier changes.

10.4 What the certificate must not do

- It must not count agents, model families, or service names as independent roots without a dependency graph.
- It must not convert a heartbeat into evidence of semantic completeness.

- It must not treat a cryptographic signature as proof that the signed content is complete or true.
- It must not omit residual ambiguity to simplify the user interface.
- It must not reuse a certificate outside its claim, authority, window, or fault budget.

11. Worked example

Suppose Enceavra maintains the claim: **No relevant provider-policy change occurred during window W.**

The primary observer has a signed checkpoint. Two research agents independently summarize the provider documentation, but both retrieve through the same cached feed. The observed vector is:

Observable	World A: no change	World B: change selectively omitted
Event receipt	none	none
Primary checkpoint	clean	clean
Observer heartbeat	ok	ok
Transport acknowledgment	ok	ok
Verifier canary	ok	ok
Agent 1 summary	no difference	no difference
Agent 2 summary	no difference	no difference

Worlds A and B are indistinguishable. The agents may use different prompts or models, but their inputs share the omission. Majority vote, debate, longer reasoning, and higher confidence cannot separate the worlds. The correct OSC decision is **UNKNOWN_WITH_RESIDUAL_AMBIGUITY**, with the residual pair recorded as **quiet** versus **issuer_selective_omission**.

Now add an external policy commitment signed from a separately operated release registry. In World A it closes clean. In World B it records the new policy version. The new channel splits the pair. Under the explicit assumption that at most one of the primary and external roots silently fails, two clean observations support **NO_RELEVANT_CHANGE**; a disagreement supports **CHANGE** or targeted revalidation.

If the external registry is populated from the same cached feed, it does not split the pair. The OSC root graph must then collapse the nominal external channel into the issuer root, and the decision returns to **UNKNOWN**.

The example shows why the mechanism is not merely "ask one more agent." The useful unit is an independently rooted observation that changes the equivalence class.

12. Brain-inspired hypothesis, kept narrow

The limited analogy is active sensing. A biological system can direct sensing actions toward information relevant to the current task [S18]. Enceavra can similarly choose a probe that maximizes the expected number or weight of target-crossing ambiguity pairs it separates.

For unresolved pair set **P** and candidate probe **a**, define the model score:

$\text{value}(a) = \sum \text{over } p \text{ in } P \text{ of } \text{weight}(p) * \text{separates}(a,p) / \text{cost}(a).$

This is a design heuristic, not a biological claim. The weights may be uniform, risk-based, or empirically estimated later. Under Round 26's evidence boundary, only uniform structural weights are used.

The testable hypothesis is:

At the same total observation cost, ambiguity-directed probes will leave fewer harmful target-crossing equivalence classes than fixed-cadence or randomly selected probes.

The necessary ablations are fixed probe order, random order, root labels removed, and a falsely independent witness. The analogy is rejected if the adaptive policy does not improve the preregistered structural and decision metrics.

13. Alternative explanations and rival designs

13.1 Conservative abstention may explain apparent safety

OSC may reduce false stability by returning UNKNOWN more often, not by learning more. E26-B must therefore report useful-current availability, false suspicion, revalidation load, and detection delay alongside harmful false-clean decisions.

13.2 An external channel may help because it is better, not because it is independent

The benefit could come from higher coverage or quality. E26-B should cross source quality with root independence where possible and include same-root high-quality and different-root matched-quality conditions.

13.3 Root metadata may be wrong

A declared witness can share infrastructure, credentials, prompts, retrieval caches, or upstream authorities. The correlation ablation shows that guarantees collapse when the graph is wrong. Operational root audits are part of the protocol, not optional documentation.

13.4 Active probes may perturb the system

Probes can create side effects, consume rate limits, reveal monitor strategy, or observe a path different from user traffic. The experiment needs synthetic, rate-limited probes, side-effect accounting, and separate coverage labels.

13.5 The finite model may omit the actual failure

An exact answer to an incomplete model can be precisely wrong. OSC must expose the model version and residual assumptions. Production decisions should include an escape path for unmodeled evidence and periodic threat-model review.

13.6 A probabilistic policy may be more useful than a zero-error certificate

Real systems may accept bounded risk. A later round can attach calibrated probabilities to equivalence classes, but probability does not remove identifiability limits. Without information or justified priors, a confident posterior can still be a restatement of assumptions.

13.7 End-to-end revalidation may dominate formal selection

Full revalidation can be simpler and safer for a small claim set. E26-B includes revalidate-all as a strong high-cost reference and compares every method at matched total verification budget.

14. Falsifiable hypotheses

ID	Hypothesis	Primary test	Rejection condition
H26.1	Under one injected silent-root fault, SACP plus one genuinely independent target channel has fewer harmful false-stability certificates than SACP at matched total cost.	Paired sandbox windows with hidden adjudication.	No improvement, or any improvement disappears after cost matching.
H26.2	Three root-disjoint target channels are sufficient against two modeled silent-root failures in the implementation.	Mutation coverage over all two-root combinations.	Any changed world is certified clean without an assumption violation.
H26.3	Removing or falsifying root-lineage metadata increases false stability under common-origin witness failures.	Independent versus mirrored witness ablation.	No difference in residual ambiguity or harmful decisions.
H26.4	Health-only checks improve abstention and reduce false certification for visible path failures but do not detect healthy-looking selective non-emission.	Heartbeat, acknowledgment, and canary ablations.	Health-only checks identify silent issuer omission without becoming target-sensitive.
H26.5	Additional LLM agents sharing the same evidence root do not match one independent external channel at equal inference and observation cost.	Same-root agent ensemble versus independent-channel policy.	Same-root ensemble matches or exceeds on false stability without regressions.
H26.6	Ambiguity-directed probes reduce residual pair weight per observation unit relative to fixed and random probes.	Active-probe ablation with frozen candidate pool.	The preregistered paired interval includes no improvement or favors baselines.
H26.7	OSC's safety gain is not fully explained by abstention.	Joint analysis of false stability and useful-current availability.	False stability falls only by exceeding the preregistered abstention or availability margin.

No result for these hypotheses is claimed in this round.

15. Proposed experiment E26-B

15.1 Prerequisites

E26-B should run only when Encevera has:

- two or three frozen current claims with explicit relevance rules;
- at least three real authority observers with recorded dependency graphs;
- hidden independent adjudication unavailable to the tested policies until verdict lock;
- an immutable replay record;
- a sandbox capable of controlled omissions without touching production data;
- approval for non-destructive, synthetic end-to-end probes.

15.2 Strata

Natural shadow windows and sandbox fault windows must remain separate.

- **Natural stratum:** 120 consecutive eligible windows, analyzed as observed without injected faults. This estimates operational abstention and workload, not rare-fault recall.
- **Sandbox stratum:** 120 replayed or synthetic windows balanced across no fault, single-root omission, two-root omission, stale or rollback, common-origin witness, verifier common mode, and negative controls. This measures fault discrimination under known ground truth.

The sample sizes are planning targets, not power claims. A preregistration should calculate detectable effects from pilot-free conservative bounds or sequential stopping rules before collection.

15.3 Policies

- 1 Receipt-absence baseline.
- 2 Round 25 SACP.
- 3 SACP plus a fixed external channel.
- 4 SACP plus OSC-selected observables.
- 5 Same-root multi-agent ensemble with the same model-call budget.
- 6 Revalidate-all high-cost reference.
- 7 Oracle using hidden adjudication, reported only as an upper reference and never as a deployable policy.

15.4 Controlled total compute and observation budget

Each non-oracle policy receives the same per-window envelope:

- at most 12 observation units;
- at most 24 model-call units if language models are used;
- the same maximum wall-clock horizon;
- the same frozen candidate observation pool;
- the same evidence retention and privacy constraints.

One observation unit is preregistered by operation class and median cost, not retrofitted after results. Unused budget is recorded. Revalidate-all may exceed the envelope and is reported as a reference, not a matched competitor.

15.5 Baselines and ablations

- Remove root labels while preserving messages.
- Relabel the witness as independent while routing it through the primary cache.
- Remove verifier canaries.
- Remove explicit clean checkpoints.
- Replace adaptive probes with fixed order.
- Replace adaptive probes with random order under identical seeds.
- Collapse all LLM agents to one model and one retrieval root.
- Vary the assumed fault budget without changing observations.

15.6 Primary metrics

Metric	Purpose
Harmful false-stability certification	Primary safety endpoint: changed window labeled NO_RELEVANT_CHANGE .
Target-crossing residual pair weight	Direct OSC endpoint before hidden truth reveal.
Change detection recall	Ability to surface covered changes.
Abstention rate	Separates safety from indiscriminate withdrawal.
False suspicion rate	Quiet windows labeled CHANGE .
Useful-current availability	Fraction of quiet windows retaining a usable scoped claim.
Detection delay	Time from adjudicated change to locked CHANGE or UNKNOWN .
Observation and model cost	Enforces the fixed-budget comparison.
Root coverage and effective root count	Detects nominal diversity without failure diversity.
Unnecessary revalidation load	Captures operational harm.

15.7 Regression and harmful-change metrics

- Privacy-bearing bytes collected by each observer.
- Probe side effects and rate-limit consumption.
- User-facing claim oscillation.
- Repeated alerts for the same unresolved equivalence class.
- Incorrect automatic restoration after a gap.
- Historical claim mutation.
- Cross-authority over-scoping.
- Decision reproducibility from the immutable evidence record.
- Verifier disagreement and unresolved verifier faults.

15.8 Analysis plan

Use paired comparisons on the same immutable windows. Lock policy outputs before hidden adjudication is revealed. Report every metric by natural versus sandbox stratum, authority family, fault root, claim, and observed root count. Do not pool injected frequency with natural frequency. Use exact paired tests or bootstrap intervals appropriate to the endpoint, publish all misses and false alerts, and report both absolute counts and denominators.

The primary superiority comparison is OSC versus SACP on harmful false-stability certification in the sandbox stratum. The co-primary noninferiority condition is useful-current availability in quiet natural windows with a preregistered margin no larger than five percentage points. Failure on either condition blocks adoption.

16. Expected failure modes of OSC

- 1 Threat-model laundering:** an unsafe world is excluded without evidence.
- 2 Root inflation:** multiple services are counted as independent despite shared infrastructure or upstream data.
- 3 Coverage overclaim:** a behavior probe is treated as covering claims it cannot exercise.
- 4 Verifier recursion:** the same component creates, evaluates, and certifies the evidence.

- 5 **Stale matrix:** system dependencies change while an old separation matrix remains valid on paper.
 - 6 **Assumption drift:** a one-fault certificate is reused after the system accepts correlated failures.
 - 7 **Availability collapse:** excessive `UNKNOWN` decisions trigger broad and costly revalidation.
 - 8 **Adversarial probe recognition:** an issuer answers canaries but omits ordinary changes.
 - 9 **Clock and expiry failure:** freshness checks share a faulty time source.
 - 10 **Evidence retention gap:** the certificate survives but raw inputs needed for independent replay do not.
 - 11 **Model incompleteness:** a real failure lies outside every enumerated world pair.
 - 12 **Metric gaming:** policies optimize ambiguity counts while ignoring harm, privacy, or user-facing stability.
- The protocol should fail closed on certificate integrity while failing scoped, not global, on target uncertainty.

17. Limitations

- The finite model is binary, atemporal within each window, and intentionally small.
- It represents worst-case silent failures but not forged positive change reports or every Byzantine interaction.
- The illustrative cost model is arbitrary and used only to demonstrate selection mechanics.
- The root-disjointness theorem depends on accurate root grouping and truthful surviving channels.
- A zero ambiguity count can reflect an incomplete world set.
- The analysis does not validate cryptography, signatures, clocks, network delivery, or source semantics.
- The multi-agent literature cited uses different tasks, models, and protocols; it motivates caution but does not validate OSC.
- Active sensing is a design analogy, not evidence of cognitive or neural equivalence.
- No independent human adjudicator reviewed E26-A because its ground truth is generated directly from the declared model.
- No live system, model, or benchmark result supports an effectiveness or deployment claim.

18. Decision and next-round handoff

18.1 Round 26 decision

Adopt OSC only as a research artifact and pre-deployment proof obligation. Do not install it as a production gate from E26-A. The operational rule carried forward is:

A quiet observation vector supports `NO_RELEVANT_CHANGE` only when its OSC equivalence class is target-pure under a declared, current, and independently reviewable threat model. Otherwise the result is `UNKNOWN`, and the public claim must be narrowed to what the evidence actually establishes.

18.2 Preferred Round 27

Run E26-B over a small real multi-authority shadow window with hidden independent adjudication, fixed budgets, and natural and sandbox strata reported separately. Publish every miss, false alert, abstention, restoration, and root-lineage correction. Do not deploy from the shadow trial.

18.3 Fallback if access remains unavailable

Do not build a larger synthetic world model. Instead, apply OSC manually to one real current claim, produce an audited root-dependency inventory, and measure whether each proposed observation would actually split a live ambiguity class. If even that access is unavailable, the next distinct theoretical question should be fixed-budget verification allocation using ambiguity reduction as the objective, with no effectiveness claim.

19. Reproducibility record

The bundle contains:

- `encevra_research_round_26.md` - manuscript source;
- `round26_identifiability.py` - exhaustive finite-model checker;
- `test_round26_identifiability.py` - deterministic tests;
- `outputs/summary.json` - model and minimum-set summary;
- `outputs/package_metrics.csv` - named policy outcomes;
- `outputs/minimal_identifying_sets.csv` - all inclusion-minimal sets;
- `outputs/ambiguity_pairs.csv` - all reported indistinguishable opposite-target pairs;
- `outputs/minimal_blinding_cuts.csv` - smallest structural cuts;
- `source_inspection.csv` - source, date, inspected material, support, and non-support ledger;
- `outputs/sha256sums.txt` - generated-output manifest;
- `run.sh`, `README.md`, and `LICENSE`.

Reproduction requires Python 3.11 or later and no third-party packages:

```
cd research/round26
./run.sh
```

The script prints minimum sets and rewrites the structured outputs deterministically. The PDF renderer is separate from the diagnostic and does not affect results.

20. Source inspection register

All web sources below were accessed on 2026-10-17. The bundle's `source_inspection.csv` records the full inspection ledger.

ID	Date	Material directly inspected	What it supports
S01-S04	2026-10-17 archive state	Research overview, rounds index, registry, full R25 and R24, relevant full R10 and R9 routes.	Numbering through Round 25, prior conclusions, and the formal fallback.
S05	1982-07	Abstract, oral-message impossibility, signed-message assumptions.	Fault and authentication assumptions determine guarantees.
S06	1985-04	Abstract and asynchronous crash model.	A missing information model can impose impossibility.
S07	1988-04	Abstract and two partial-synchrony definitions.	Added timing assumptions change solvability.
S08	1990-07	Abstract, indistinguishability relation, temporal imprecision, weaker knowledge variants.	Knowledge is relative to possible points with the same view.
S09	1995-09	IEEE abstract and diagnosability definition.	Fault inference depends on observable sequences.
S10	1996-03	Abstract, model, completeness and accuracy, weakest detector discussion.	Failure detectors provide explicitly characterized information.
S11	1998-07	IEEE abstract and active-diagnosis framing.	Actions can be chosen to disambiguate faults.

ID	Date	Material directly inspected	What it supports
S12	2007-10-14	Abstract, assumptions, witness configuration, evaluation boundary.	Accountability is conditional on observed effects and correct witnesses.
S13	2021-12	Introduction, timestamps, inclusion, consistency, monitoring limits.	Append-only commitments aid auditing but require monitors.
S14	current spec retrieved 2026-10-17	Threat model, freeze and rollback rules, expiry, thresholds.	Versions and freshness convert some stale states into detectable failures.
S15	ICML 2024	Abstract, results overview, agents and rounds discussion.	Multi-agent debate can improve selected benchmark outcomes.
S16	2025-09-05	Abstract, setup, result table, discussion.	Debate can also degrade accuracy in tested settings.
S17	2025-03-17	Dataset, taxonomy, verification and termination failures, limitations.	Current multi-agent systems exhibit structural coordination and verification failures.
S18	2016-10	Observer-planner loop, value of information, stopping, evidence caution.	Limited active-sensing design analogy.

21. Full sources

[S01] Encevera Research. Research overview, rounds index, and registry. Lovable project [8483bc59-da36-492e-b8b8-6bf684070241](#), inspected at commit [f1197fa663970737c3da35bc8e98c63c60f59c78](#), 17 October 2026.

[S02] Encevera Research. "Round 25: When Silence Looks Like Stability: Detecting Missing Change Evidence in Collaborative AI Systems." 16 October 2026. Full manuscript inspected in the same Lovable project.

[S03] Encevera Research. "Round 24." 15 October 2026. Full manuscript inspected in the same Lovable project.

[S04] Encevera Research. "Round 10: Evidence exchange, selective revision, and a testable path toward brain-inspired intelligence" and Round 9 on learning reusable procedures from verified experience. Full route texts inspected in the same Lovable project, 17 October 2026.

[S05] Lamport, Leslie, Robert Shostak, and Marshall Pease. "The Byzantine Generals Problem." ACM Transactions on Programming Languages and Systems 4, no. 3 (July 1982): 382-401. [PDF](#). DOI: [10.1145/357172.357176](#).

[S06] Fischer, Michael J., Nancy A. Lynch, and Michael S. Paterson. "Impossibility of Distributed Consensus with One Faulty Process." Journal of the ACM 32, no. 2 (April 1985): 374-382. [ACM record](#). DOI: [10.1145/3149.214121](#).

[S07] Dwork, Cynthia, Nancy Lynch, and Larry Stockmeyer. "Consensus in the Presence of Partial Synchrony." Journal of the ACM 35, no. 2 (April 1988): 288-323. [PDF](#). DOI: [10.1145/42282.42283](#).

[S08] Halpern, Joseph Y., and Yoram Moses. "Knowledge and Common Knowledge in a Distributed Environment." Journal of the ACM 37, no. 3 (July 1990): 549-587. [PDF](#). DOI: [10.1145/79147.79161](#).

[S09] Sampath, Meera, Raja Sengupta, Stephane Lafortune, Kasim Sinnamohideen, and Demosthenis Teneketzis. "Diagnosability of Discrete-Event Systems." IEEE Transactions on Automatic Control 40, no. 9 (September 1995): 1555-1575. [IEEE record](#). DOI: [10.1109/9.412626](#).

[S10] Chandra, Tushar Deepak, and Sam Toueg. "Unreliable Failure Detectors for Reliable Distributed Systems." Journal of the ACM 43, no. 2 (March 1996): 225-267. [PDF](#). DOI: [10.1145/226643.226647](#).

[S11] Sampath, Meera, Stephane Lafortune, and Demosthenis Teneketzis. "Active Diagnosis of Discrete-Event Systems." IEEE Transactions on Automatic Control 43, no. 7 (July 1998): 908-929. [IEEE PDF](#). DOI: [10.1109/9.701089](#).

[S12] Haeberlen, Andreas, Petr Kouznetsov, and Peter Druschel. "PeerReview: Practical Accountability for Distributed Systems." Proceedings of the 21st ACM Symposium on Operating Systems Principles, 2007. [PDF](#). ACM record: [10.1145/1323293.1294279](#).

[S13] Laurie, Ben, Eran Messeri, and Rob Stradling. "Certificate Transparency Version 2.0." RFC 9162, December 2021. [RFC Editor](#). DOI: [10.17487/RFC9162](#).

[S14] The Update Framework. "The Update Framework Specification," current 1.x specification, retrieved 17 October 2026. [Specification](#).

[S15] Du, Yilun, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. "Improving Factuality and Reasoning in Language Models through Multiagent Debate." ICML 2024. [Project and paper](#). [OpenReview](#).

[S16] Wynn, Andrea, Harsh Satija, and Gillian Hadfield. "Talk Is Not Always Cheap: Understanding Failure Modes in Multi-Agent Debate." arXiv:2509.05396, 5 September 2025. [arXiv](#).

[S17] Cemri, Mert, et al. "Why Do Multi-Agent LLM Systems Fail?" arXiv:2503.13657, 17 March 2025. [arXiv](#). [OpenReview](#).

[S18] Yang, Scott Cheng-Hsin, Daniel M. Wolpert, and Mate Lengyel. "Theoretical Perspectives on Active Sensing." Current Opinion in Behavioral Sciences 11 (October 2016): 100-108; author manuscript available in PMC in 2018. [PMC](#). DOI: [10.1016/j.cobeha.2016.06.009](#).

22. Closing statement

Round 25 taught Enceavra to distrust unexplained silence. Round 26 makes the stronger boundary explicit: some silence is not merely hard to detect; it is indistinguishable from stability under the available view. Reliability improves when the system records that fact, buys an observation that actually separates the worlds, or narrows its claim. It does not improve merely because more agents discuss the same missing evidence.