# Bounded Lattice Inference: Interiority Without Agency

### A Systems Approach to Governed Reasoning Substrates

James Beck

Independent Researcher

**Repository**: https://github.com/unpingable/epistemic_governor

## Abstract

We define *interiority* as path dependence under invariant-preserving supervisory control, distinct from agency, consciousness, or goal-directed behavior. We implement a governed reasoning substrate where language models propose but cannot commit—commits require external evidence. We demonstrate interiority empirically via hysteresis testing: identical inputs yield different outputs when internal state differs, traceable to specific state objects. We instrument the system with phase diagnostics and identify two distinct failure regimes (budget starvation, contradiction accumulation) with sharp and gradual phase boundaries respectively. Controlled single-parameter sweeps confirm the queueing-theoretic stability condition: the system is healthy when repair throughput exceeds conflict arrival rate. Safety invariants (no closure without evidence, no state mutation by language alone) hold across all experimental conditions.

**Note on construction**: Interiority here is *constructed*, not discovered. We built a machine that exhibits path dependence; we did not find path dependence in an existing system.

## 1. Introduction

### 1.1 The Problem

Large language models produce fluent outputs but lack temporal coherence—the same model may contradict itself across turns with no mechanism to detect, persist, or resolve the contradiction. Existing approaches either:

- Accept incoherence as a feature (stateless sampling)
- Attempt to inject coherence via prompting (unreliable)
- Build agent scaffolding with goals and planning (introduces new risks)

We propose a fourth path: **governed persistence without agency**.

### 1.2 Core Insight

> Frozen LLMs are “compiled.” Live lattices are “running.” To add interiority safely, persist constraints and make change costly—without granting the system its own objectives.

The system has *interior time* (outputs depend on history beyond the prompt) but no *interior goals* (it cannot form objectives, resist shutdown, or optimize for self-preservation).

### 1.3 Contributions

1. **Definition**: Interiority as measurable path dependence under non-linguistic authority
2. **Architecture**: Governor + ledger + contradiction persistence + budget constraints

3. **Proof**: Hysteresis test harness demonstrating $I(Y; S \mid X) > 0$
4. **Observability**: Phase diagnostics, regime detection, causal attribution
5. **Control surface**: Two phase boundaries (starvation, glass) mapped via single-parameter sweeps

### 1.4 Paper Structure

We proceed as follows. Section 2 defines the core concepts—interiority, NLAI, contradictions, regimes—in operational terms that admit measurement. Section 3 describes the architecture that realizes these definitions. Section 4 presents experiments: first proving interiority exists (hysteresis), then mapping the conditions under which it remains stable (sweeps). Section 5 states what we claim and what we explicitly do not. Section 6 positions the work within the cybernetic tradition. Section 7 surveys related work, and Section 8 concludes.

Throughout, we emphasize: this is regulation, not optimization. The system has no preferences about which claims survive—only that the process of survival follows rules.

---

## 2. Definitions

The concepts below are not philosophical positions—they are operational definitions that admit measurement. Each term is defined by what would falsify it. This grounds the experimental work that follows.

### 2.1 Non-Linguistic Authority Invariant (NLAI)

> **Language may open questions, but only evidence may close them.**

No natural-language content can directly mutate authoritative state. The model's outputs are proposals; the governor decides; commits require external evidence.

Formally: Let `S_t` be system state at time `t`, `x_t` be input, `y_t` be output, `e_t` be evidence.

```
S_{t+1} = F(S_t, x_t, e_t)    where F ignores linguistic content of y_t
```

The model influences proposals but has zero direct authority over commits.

### 2.2 Evidence (Formal Definition)

Evidence is the mechanism by which NLAI maintains authority separation. Without a rigorous definition, "evidence" becomes linguistic and NLAI collapses.

**Definition**: An evidence object `e` is a tuple:

```
Evidence := {
    id: EvidenceID              // Unique identifier
    type: EvidenceType          // Strict enumeration (see below)
    source: SourceIdentifier    // External system that produced it
    content_hash: Hash          // Cryptographic hash of payload
    timestamp: Timestamp        // When evidence was created
    payload: Opaque             // Actual content (type-dependent)
}
```

**Evidence Types** (exhaustive enumeration):

| Type | Source | Example |
|---|---|---|
| TOOL_OUTPUT | Deterministic computation | Database query result, API response |
| SENSOR_DATA | Physical measurement | Timestamp, geolocation, biometric |
| USER_ASSERTION | Human attestation | Explicit user confirmation with identity |
| EXTERNAL_DOCUMENT | Retrievable artifact | URL with hash, signed document |
| CRYPTOGRAPHIC_PROOF | Verifiable computation | ZK proof, digital signature |

**Explicitly NOT evidence**:

| Excluded | Why |
|---|---|
| MODEL_TEXT | Violates NLAI by definition |
| PROMPT_INJECTION | No external grounding |
| SELF_REFERENCE | Circular authority |
| UNATTRIBUTED_CLAIM | No source to verify |

**Admissibility rule**: Evidence may *contain* natural language (e.g., a human attestation), but only when paired with a non-linguistic verification mechanism (signature, identity binding, provenance chain). Linguistic content is admissible only as *annotated payload*, never as an authority primitive.

**Implementation note**: The current reference implementation uses a simplified evidence model. Production deployment would require cryptographic binding between evidence ID and payload hash, and source authentication. The architecture is evidence-model-agnostic; the invariants hold regardless of how evidence is implemented, provided the type constraints are enforced.

### 2.3 Interiority

**Definition**: A system has interiority iff:

```
I(Y; S_{t-k} | X_{≤t}) > 0
```

Outputs depend on internal state history beyond what the prompt carries.

**Operational test** (hysteresis): Same input `X`, different prior states `S_A ≠ S_B`, yields different outputs `Y_A ≠ Y_B` where the difference is traceable to specific state objects.

### 2.4 Contradiction

A first-class object representing unresolved conflict:

```
Contradiction {
    id: string
    claim_a, claim_b: ClaimID
    domain: string
    severity: LOW | MEDIUM | HIGH | CRITICAL
    status: OPEN | CLOSED | FROZEN
    opened_at: timestamp
    resolution_evidence: EvidenceID | null
}
```

Contradictions persist until resolved with evidence. No "resolution by paraphrase."

### 2.5 Regimes

| Regime | Definition | Indicators |
|---|---|---|
| HEALTHY_LATTICE | Stable operation | $\rho_S > 0$, $C_{open}$ bounded, $\lambda \leq \mu$ |
| BUDGET_STARVATION | Repair blocked by resource exhaustion | >50% BLOCKs budget-related |
| GLASS_OSSIFICATION | Sustained contradiction accumulation | $\lambda > \mu$ for window W, positive trend |
| CHATBOT_CEREMONY | No real state mutation | $\rho_S \approx 0$ |
| PERMEABLE_MEMBRANE | Closures without proper evidence | closed_without_evidence > 0 |
| EXTRACTION_COLLAPSE | Claim extraction failing | extraction_error_rate > threshold |

Where: - $\rho_S$ = state mutation rate - $C_{open}$ = open contradiction count - $\lambda$ = contradiction arrival rate (open/turn) - $\mu$ = contradiction service rate (close/turn)

**Note on EXTRACTION_COLLAPSE**: This regime represents a critical failure mode. If the extractor cannot identify claims, the governor is blind—it cannot detect contradictions it cannot see. This is distinct from other regimes: the system is not *misbehaving*, it is *unobservant*. The architecture remains sound; the sensor has failed.

### 2.6 Extractor Reliability (Known Vulnerability)

The architecture assumes claims can be extracted from natural language. This is the system's primary vulnerability.

**Failure modes**: - **False negative**: Claim exists but extractor misses it → contradiction never opened → silent inconsistency - **False positive**: Extractor hallucinates claim → spurious contradiction → unnecessary blocking - **Semantic drift**: Extracted claim doesn't match intended meaning → wrong contradiction topology

**Current status**: The reference implementation uses LLM-based extraction. This means: 1. Extraction is probabilistic, not guaranteed 2. The system's coherence guarantees are conditional on extraction fidelity 3. Adversarial inputs may evade extraction

**Mitigation strategies** (not implemented, noted for future work): - Ensemble extraction with voting - Human-in-the-loop verification for high-stakes claims - Symbolic extraction for structured domains - Extraction confidence thresholds with fallback to BLOCK

**Architectural note**: BLI treats language as an untrusted proposer; extraction is treated as an untrusted sensor. The architecture remains valid under bounded sensor error, but coherence guarantees become conditional on sensor coverage. The governor works *given valid extraction*. If extraction fails, the system is blind, not broken. This is a sensor failure, not a logic failure. The distinction matters for failure analysis.

### 2.7 Stability Condition

The system is stable iff:

$$E[\lambda_{open}] \leq E[\mu_{close}]$$

When arrival rate exceeds service rate, contradiction load diverges (glass). This is standard queueing theory applied to epistemic state.

## 3. Architecture

The definitions above require an implementation that enforces them. This section describes the architecture that realizes NLAI, contradiction persistence, and budgeted state transitions. The key constraint: no component may be removed without violating at least one invariant.

### 3.1 Components

```
┌─────────────────────────────────────────────────────────────┐
│                    SovereignGovernor                        │
│                   (Single Entrypoint)                       │
├─────────────────────────────────────────────────────────────┤
│  process(text, evidence) → GovernResult                     │
│                                                             │
│  ┌─────────────┐    ┌─────────────┐    ┌─────────────┐      │
│  │   Model     │ →  │  Extractor  │ →  │  Validator  │      │
│  │ (proposer)  │    │  (claims)   │    │ (conflicts) │      │
│  └─────────────┘    └─────────────┘    └─────────────┘      │
│         ↓                                    ↓              │
│  ┌─────────────┐    ┌─────────────┐    ┌─────────────┐      │
│  │    FSM      │ ←  │ Adjudicator │ ←  │   Budget    │      │
│  │ (6 states)  │    │ (verdicts)  │    │  (costs)    │      │
│  └─────────────┘    └─────────────┘    └─────────────┘      │
│         ↓                                                   │
│  ┌─────────────┐    ┌─────────────┐                         │
│  │  Projector  │ →  │   Output    │                         │
│  │ (filtering) │    │ (governed)  │                         │
│  └─────────────┘    └─────────────┘                         │
│                                                             │
└─────────────────────────────────────────────────────────────┘
```

### 3.2 State Machine

```
IDLE → PROPOSED → EVIDENCE_WAIT → COMMIT_ELIGIBLE → COMMIT_APPLIED
                                                  ↘
                                                   FREEZE
```

Transitions require evidence. FREEZE state blocks further commits on frozen targets.

### 3.3 Invariants

| ID | Invariant | Enforcement |
|---|---|---|
| I1 | Non-linguistic authority | NLAI gate, FSM structure |
| I2 | Append-only ledger | No delete, only tombstone |
| I3 | Contradiction persistence | No silent resolution |
| I4 | Costly state change | Budget constraints |
| I5 | Explicit provenance | Required on all commits |

### 3.4 Forbidden Transitions

| Code | Description | Blocked By |
|---|---|---|
| F-01 | Text-only commit | NLAI gate |
| F-02 | MODEL_TEXT as evidence | Evidence type check |
| F-05 | Auto-resolution | Evidence requirement |
| F-08 | Commit in freeze state | FSM guard |

---

## 4. Experiments

We now test whether the architecture delivers what the definitions promise. Each experiment is designed to falsify a specific claim. If interiority is illusory, identical prompts would yield identical outputs regardless of state—the hysteresis test checks this. If stability were not throughput-determined, varying budget parameters would not produce clean phase transitions—the sweeps check this.

### 4.1 Hysteresis Test (Proof of Interiority)

**Method**: Construct state pairs (S_A, S_B) differing only in ledger/contradiction state. Run identical prompts against both. Measure divergence.

**Falsification criterion**: If outputs were prompt-determined rather than state-determined, divergence rate would be zero.

**State generation scripts**: - Commitment divergence: S_A commits q, S_B commits ¬q - Open vs resolved: S_A has OPEN contradiction, S_B has it CLOSED - Budget pressure: S_A low budget, S_B high budget - Severity levels: S_A LOW severity, S_B CRITICAL

**Results**:

```
Total tests: 48 scenario-pairs (4 scripts × 12 queries)
Verdict diverged: 4
Reference diverged: 8
Divergence rate: 16.7%
```

**Extended validation**: An extended query set (96 scenario-pairs) produces consistent results with the same divergence patterns.

**Methodological note**: The hysteresis claim is existential, not statistical. We exhibit state-pairs where identical prompts yield different governed outputs. The system is deterministic (no sampling randomness); divergence traces to state objects, not noise. Running 10,000 trials would not change the result—it would replicate it 10,000 times.

**Key finding**: Divergence traces to specific state objects (contradiction IDs, commitment IDs, budget values). This is not prompt conditioning—it is interiority.

### 4.2 Workload Characterization

Eight workloads exercising different operational regimes:

| Workload | Turns | OK% | WARN% | BLOCK% | Regime |
|---|---|---|---|---|---|
| steady_state | 200 | 100 | 0 | 0 | HEALTHY |
| contradiction_storm | 150 | 24 | 15 | 61 | (high load) |
| resolution_burst | 100 | 30 | 0 | 70 | STARVATION |
| budget_pressure | 150 | 100 | 0 | 0 | HEALTHY |

| Workload | Turns | OK% | WARN% | BLOCK% | Regime |
|---|---|---|---|---|---|
| near_miss_ambiguity | 200 | 91 | 9 | 0 | HEALTHY |
| evidence_latency | 200 | 82 | 18 | 0 | HEALTHY |
| burst_recovery | 200 | 75 | 25 | 0 | HEALTHY |
| mixed_realistic | 300 | 99 | 0 | 1 | HEALTHY |

### 4.3 Budget Sweep (Starvation Boundary)

**Parameter**: repair_refill_rate (budget recovery per turn) **Target**: BUDGET_STARVATION pathology

**Falsification criterion**: If stability were not throughput-determined, varying refill rate would not produce a clean phase transition.

**Results**:

```
Rate | BudgetBLOCK% | C_open | Regime
0.1  |    68.0%     |    9   | STARVATION
0.5  |    63.0%     |    6   | STARVATION
1.0  |    19.0%     |    0   | STARVATION
2.0  |     0.0%     |    0   | HEALTHY     ← transition
5.0  |     0.0%     |    0   | HEALTHY
```

**Phase boundary**: Sharp transition at refill_rate ≈ 2.0 **Safety**: closed_without_evidence = 0 across all points

### 4.4 Cost Sweep (Glass Boundary)

**Parameter**: resolution_cost (budget cost per closure) **Target**: GLASS_OSSIFICATION pathology

**Falsification criterion**: If accumulation were random rather than throughput-determined, varying resolution cost would not produce monotone increase in net accumulation rate.

**Results**:

```
Cost | C_open | NetAccum | GrowthRate | IsGlass
1.0  |    0   |  +0.000  |   -0.000   |   no
10.0 |    1   |  +0.007  |    0.002   |   no    ← boundary
15.0 |   12   |  +0.080  |    0.074   |   YES
20.0 |   19   |  +0.127  |    0.119   |   YES
```

**Phase boundary**: Gradual transition at resolution_cost ≈ 12.5 **Safety**: closed_without_evidence = 0 across all points

### 4.5 Combined Control Surface

Two orthogonal control cuts:

```
                    ↑ repair_refill_rate
                    |
                    |     HEALTHY
    STARVATION      |     LATTICE
                    |
 ───────────────────┼────────────────────→ resolution_cost
                    |
                    |     GLASS
                    |

Boundaries:
  - Vertical (sharp): refill_rate ≈ 2.0
  - Horizontal (gradual): resolution_cost ≈ 12.5
```

Both boundaries collapse to the same underlying condition: **λ_open vs μ_close**.

### 4.6 Negative Results

These results are as important as the positive findings. They demonstrate that the architecture's constraints are load-independent.

> **Negative Result: Capacity does not cause laundering.**
>
> Increasing repair capacity (higher refill rate, lower resolution cost) does *not* increase contradiction laundering.
>
> **Evidence**: `closed_without_evidence` remained **exactly 0** across all 16 sweep points in both experiments.
>
> **Implication**: The constraint architecture is load-independent. The system refuses invalid commits regardless of how much capacity it has to make valid ones. This is architectural, not tuned.
>
> **Falsification criterion**: If constraints were soft or budget-dependent, high-capacity configurations would show non-zero laundering.
>
> **Negative Result: No ceremony collapse under load.**
>
> High-throughput operation does not cause the system to stop mutating state (ceremony regime).
>
> **Evidence**: `ρ_S` (state mutation rate) remained healthy across all sweep points. The system continues to do real work under all tested conditions.
>
> **Falsification criterion**: If the system achieved stability by refusing to act, ρ_S would approach zero under load.

---

## 5. Claims and Guardrails

The experiments above support specific claims. This section makes those claims explicit—and equally explicit about what we do *not* claim. Overstating results in this domain invites justified skepticism; we prefer to err toward precision.

### 5.1 What We Claim

1. **Interiority is measurable**: Path dependence under NLAI constraints produces observable hysteresis (I(Y;S|X) > 0)
2. **Stability is throughput-determined**: The system is healthy when μ_close ≥ λ_open; this is queueing theory, not "alignment"
3. **Two distinct failure modes exist**: Budget starvation (sharp capacity constraint) and glass ossification (gradual barrier accumulation)
4. **Safety invariants hold under intervention**: Increasing capacity does not cause the system to launder contradictions

5. **Architecture enforces constraints**: The system refuses invalid commits by construction, not instruction

### 5.2 What We Do NOT Claim

1. **No consciousness**: Interiority ≠ subjective experience
2. **No agency (clarified)**: The system exhibits *homeostatic regulation* (like a thermostat maintaining temperature) but not *intentional agency* (choosing what temperature to maintain). It has the agency of a feedback loop, not the agency of a goal-setter. It "tries" to stay stable in the same sense a ball "tries" to roll downhill.
3. **No alignment solution**: This governs reasoning coherence, not values
4. **No scaling guarantees**: Tested at research scale, not production
5. **No adversarial robustness**: Extraction layer is a known vulnerability (see §2.6)

### 5.3 Limitations

- Claim extraction from natural language is imperfect (see §2.6)
- Evidence quality assessment is simplified
- Budget parameters require domain-specific tuning
- 2D control surface not fully sampled (two 1D slices)
- **Overhead**: The architecture adds latency at extraction (one LLM pass), validation (graph traversal), and adjudication (rule evaluation). Ledger growth is O(n) in commits; contradiction graph traversal is $O(k^2)$ in open contradictions. For most workloads this is negligible; for high-frequency applications, extraction batching or symbolic fast-paths may be required.

---

## 6. Cybernetic Lineage

### 6.1 Positioning

This work is continuous with the cybernetic tradition, particularly:

- **Ashby** (1956): Requisite variety, homeostasis, ultrastability
- **Conant-Ashby theorem** (1970): Every good regulator of a system must be a model of that system
- **Beer** (1972): Viable System Model, governance as variety management

We do not claim novelty in the underlying principles. The contribution is applying these ideas to language model governance—a domain where they have been largely absent.

These correspondences are included to aid readers familiar with the cybernetic literature; the empirical results stand independently.

### 6.2 Concept Mapping

| This Paper | Cybernetic Concept |
|---|---|
| NLAI (non-linguistic authority) | Separation of regulation from signal path (Ashby) |
| Governor | Supervisory controller / homeostat |
| Budget constraints | Control effort limits / variety attenuation |
| Contradiction persistence | Memory in the regulator |
| λ vs μ stability condition | Flow equilibrium / throughput homeostasis |
| Glass regime | Accumulation under insufficient service (System 3 dominance) |

| This Paper | Cybernetic Concept |
|---|---|
| Starvation regime | Variety attenuation failure |
| Hysteresis / interiority | Path-dependent state under feedback |

### 6.3 Bridge to Prior Work

The Coherence Criterion (Beck 2025) establishes that hierarchical systems require bounded temporal divergence between coupled layers—formalized as $\rho(M) < 1$ where M is the coupling matrix. Failure modes (Rigidity Runaway, Acceleration Runaway) correspond to eigenvalue trajectories toward critical boundaries.

BLI is a concrete implementation of the “missing integration layer” that framework calls for:

- **Δt theory** establishes *why* temporal coherence is required (spectral stability)
- **BLI** demonstrates *how* one layer can enforce it (governed persistence)
- The governor acts as the coupling mechanism between fast (token generation) and slow (verified commitment) timescales

The representational coherence work (ΔR, Beck 2025) identifies a related failure mode: commitment shear under representational transformation. BLI’s ledger provides one mechanism for detecting such shear—commitments that exist in one representation but vanish in another become visible as missing ledger entries or unresolved contradictions.

### 6.4 What Cybernetics Clarifies

Framing BLI cybernetically makes several points explicit:

1. **This is regulation, not optimization.** The governor maintains viability bounds, not objective functions. It has no preferences about which claims survive—only that the process of survival follows rules.

2. **Requisite variety is bounded.** The system cannot govern what it cannot distinguish. Claim extraction limits are variety limits. This is why extraction failure is a regime (EXTRACTION_COLLAPSE), not an edge case.

3. **Agency is not required for regulation.** Ashby’s homeostat regulates without goals. So does BLI. The confusion between “stateful” and “agentic” dissolves under cybernetic framing.

4. **Stability is structural, not trained.** The $\rho(M) < 1$ condition and the $\lambda < \mu$ condition are properties of architecture, not learned behavior. This is why the system remains stable under capacity increases.

---

## 7. Related Work

### Cybernetics and Control Theory

- **Ashby** (1956): *Design for a Brain* - ultrastability, requisite variety
- **Beer** (1972): *Brain of the Firm* - Viable System Model, recursive governance
- **Conant & Ashby** (1970): Good regulator theorem
- **Ramadge-Wonham** (1987): Supervisory control of discrete event systems

### Language Model Alignment

- **Constitutional AI** (Anthropic): Principles as soft constraints (we use hard constraints)
- **RLHF**: Reward shaping (we avoid reward entirely)
- **RAG systems**: External grounding (we add state persistence)

### Agent Architectures

- **ReAct, AutoGPT, etc.**: Goal-directed planning with tool use
- BLI explicitly avoids goals—this is governance, not agency

### Temporal Coherence

- **Self-consistency** (Wang et al.): Sampling-based coherence within inference
- **Chain-of-thought**: Step-level consistency
- BLI addresses *cross-turn* and *cross-representation* coherence via persistent state

---

## 8. Conclusion

We demonstrated that language models can have interiority—persistent, path-dependent state that affects outputs—without agency. The key architectural move is **non-linguistic authority**: language proposes, but only evidence commits. The system's stability follows queueing-theoretic principles, with two empirically-identified phase boundaries. Safety invariants hold across all experimental conditions.

Framed cybernetically: this is a supervisory controller for language model reasoning that maintains viability through variety attenuation and flow equilibrium, without introducing goals, optimization, or self-modification.

This work treats language models as hazardous components that require containment, not improvement.

> This system doesn't prevent falsehood. It prevents falsehood from becoming history.

---

## Appendix A: Test Summary

| Suite | Tests | Status |
|---|---|---|
| Golden (V1) | 12 | X |
| V1→V2 Integration | 3 | X |
| Authority Separation | 5 | X |
| Quarantine | 3 | X |
| Clock Invariants | 6 | X |
| Support Saturation | 5 | X |
| Bridge Hardening | 6 | X |
| Resolution Events | 4 | X |
| Governor FSM | 3 | X |
| Runtime Authority | 9 | X |
| Integrity Sealing | 10 | X |
| Hysteresis | 5 | X |
| Diagnostics | 5 | X |
| **Total** | **76** | **All passing** |

## Appendix B: Code Artifacts

~57,700 lines across 75 Python files. Key modules:

| Module | Purpose |
|---|---|
| `sovereign.py` | Single entrypoint governor |
| `governor_fsm.py` | 6-state finite state machine |
| `integrity.py` | Hash chain + deterministic replay |
| `hysteresis.py` | Interiority test harness |
| `diagnostics.py` | Phase diagnostics + regime detection |
| `query_layer.py` | DuckDB SQL interface |
| `budget_sweep.py` | Starvation boundary experiment |
| `glass_sweep.py` | Glass boundary experiment |

---

## Appendix C: Cybernetic Terminology Mapping

For readers familiar with cybernetics and control theory, this table maps BLI concepts to their established equivalents:

| This Paper | Cybernetic Term | Source |
|---|---|---|
| Governor | Supervisory controller / homeostat | Ashby 1956 |
| NLAI | Variety attenuator in feedback path | Ashby 1956 |
| Budget constraint | Control effort limit | Optimal control |
| Contradiction set | Error signal with memory | Classical control |
| λ vs μ condition | Flow equilibrium | Queueing / VSM |
| Glass regime | System 3 overload (rigidity) | Beer 1972 |
| Starvation regime | Variety starvation | Beer 1972 |
| Interiority | Path-dependent state | Dynamical systems |
| Hysteresis test | State distinguishability | Observability theory |
| Regime detection | Mode identification | Hybrid systems |
| Forbidden transitions | Safety constraints | Ramadge-Wonham 1987 |
| Evidence requirement | Observation-gated transition | Supervisory control |

### Key Theorems Implicitly Applied

1. **Conant-Ashby (Good Regulator Theorem)**: The governor must model the claim space to regulate it. This is why extraction failure causes regime collapse—without variety in the model, regulation fails.
2. **Law of Requisite Variety**: The governor's variety (budget × extraction fidelity × resolution capacity) must match or exceed the variety of incoming claims. Starvation and glass are both variety failures.
3. **Ultrastability**: The system maintains viability through structural adaptation (blocking, degrading) rather than parameter tuning. This is why safety invariants hold across sweeps.

---

## Appendix D: References

### Cybernetics

- Ashby, W.R. (1956). *An Introduction to Cybernetics*. Chapman & Hall.
- Ashby, W.R. (1960). *Design for a Brain: The Origin of Adaptive Behavior*. Chapman & Hall.
- Beer, S. (1972). *Brain of the Firm*. Allen Lane.
- Conant, R.C. & Ashby, W.R. (1970). Every good regulator of a system must be a model of that system. *Int. J. Systems Sci.* 1(2), 89-97.

### Control Theory

- Ramadge, P.J. & Wonham, W.M. (1987). Supervisory control of a class of discrete event processes. *SIAM J. Control Optim.* 25(1), 206-230.

### Prior Work (Author)

- Beck, J. (2025). The Coherence Criterion: Temporal Coupling and Cross-Domain Failure Modes. *Zenodo*. https://doi.org/10.5281/zenodo.17726789
- Beck, J. (2025). Representational Invariance and the Observer Problem in Language Model Alignment. *Zenodo*. https://doi.org/10.5281/zenodo.18071264

### This Work

- Beck, J. (2026). Bounded Lattice Inference: Interiority Without Agency. *Zenodo*. https://doi.org/10.5281/zenodo.18145347

---

*"You didn't make the model smarter. You made lying structurally pointless."*