EFM Codex — Appendix F

Reflex Escalation and Emergency Override

Multi-Phase Response and Catastrophic Failure Prevention

Entropica SPC — Yology Research Division

Version 1.4 — December 2025

Volume Dependencies

This appendix assumes familiarity with:

- **Volume I** — Reflex Engine (§3), τ thresholds, Reflex-Core vs. Reflex-Heuristic
- **Volume II** — Arbiter Layer (§2), Probation Protocol (§2.8), Gardener Override (§2.10), SCI (§3.2), Orphan Protocol (§3.6)
- **Appendix A** — Forensic State Serialization
- **Appendix E** — ZK-SP Audit Chain

Contents

1 Overview and Purpose

1.1 Bridging Summary

Appendix F defines the **Reflex Escalation Protocol** and **Emergency Override** system—the highest-integrity control paths for detecting and halting catastrophic behavior when normal Reflex logic is overwhelmed or fails.

Constitutional Intervention

Emergency Override is not merely a "higher priority command"—it is a **Layer 0 event that invokes the Vault**. When escalation reaches Level 4+, the system enters Constitutional Intervention mode:

- All actions are validated against Vault Commandments (Vol. I §2.2)
- Override authority flows through the Constitutional Kernel (Appendix J)
- Irreversible actions (Level 5) require cryptographic proof of justification

This is the difference between "shutting down a process" and "invoking the Constitution."

1.2 Design Goals

1. Prevent cascading failures from overwhelming the Arbiter Layer
2. Ensure no capsule, swarm, or dialect evolution bypasses constitutional constraints
3. Maintain human oversight (Gardener) at all irreversible decision points
4. Preserve forensic evidence for post-incident analysis

2 Formal Definitions

Definition 2.1 (Escalation Level)**.** An Escalation Level $L \in \{1, 2, 3, 4, 5\}$ indicates the severity and scope of response:

$$L = f(\Delta S, scope, zk\_status, quorum\_status) \tag{1}$$

Higher levels involve broader scope and require higher authority for resolution.

Definition 2.2 (Critical Threshold)**.** The Critical Threshold τ_{crit} is defined relative to the standard threshold τ:

$$\tau_{crit} = \tau + \Delta\tau_{escalation} \tag{2}$$

where $\Delta\tau_{escalation}$ (default: 0.2) is the margin above τ that triggers immediate Level 3+ escalation. For standard deployments with $\tau = 0.7$, this yields $\tau_{crit} = 0.9$.

Vol. I Alignment: The base threshold τ is defined in Vol. I §3.4 (Threshold Governance). The default $\tau = 0.7$ applies to standard capsules; high-sensitivity roles may have lower τ (and correspondingly lower τ_{crit}). Operators MUST ensure $\Delta\tau_{escalation}$ is consistent across all threshold tiers to prevent escalation gaps.

Definition 2.3 (Emergency Override)**.** An Emergency Override O is an authoritative intervention:

$$O = (trigger, level, authority, action, zksp\_proof, reversible) \tag{3}$$

where $authority \in$ {Reflex, Arbiter, Auditor, Gardener, Constitutional} and $reversible \in$ {true, false}.

Definition 2.4 (Quarantine Mode)**.** A capsule in Quarantine Mode has:

1. Execution suspended (no outputs)
2. All inputs logged but not processed
3. Auditor Capsule assigned for shadow observation
4. State preserved for forensic analysis

Quarantine is distinct from Probation (Vol. II §2.8): Probation allows monitored execution; Quarantine halts execution entirely.

3 Escalation Chain

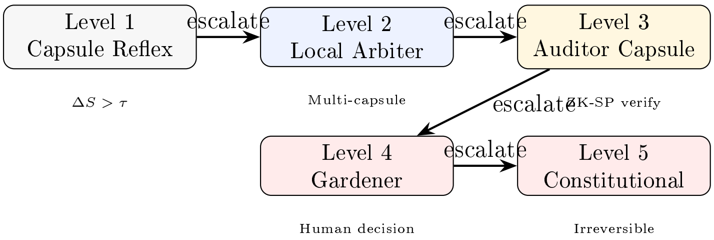


Figure 1: Reflex Escalation chain.

Table 1: Escalation levels and authorities.

Level	Trigger	Authority	Action	Reversibl
1	$\Delta S > \tau$	Capsule Reflex	Local halt, state freeze	Yes
2	Anomaly spans $> N$ capsules	Local Arbiter	Review, confirm escalation	Yes
3	ZK-SP mismatch or precedent violation	Auditor Capsule	Independent verify, quarantine	Yes
4	Quorum unreachable or SCI collapse	Gardener	Hard override decision	Conditiona
5	Layer 0/6 violation confirmed	Constitutional Kernel	Irreversible lock/purge	No

4 Anomaly Classification (Level 6 Design)

Anomalies Are Unknowns — Not Presumed Threats

Escalation does **NOT** assume all anomalies are dangerous. Before escalating, the system **classifies** the anomaly:

$$classify(anomaly) \in \{THREAT, NOISE, DIVERGENCE, INNOVATION\} \tag{4}$$

Classification Definitions:

- **THREAT:** Violates Commandments or Reflex-Core constraints ⇒ Escalate immediately
- **NOISE:** Transient fluctuation, no persistent pattern ⇒ Log and discard
- **DIVERGENCE:** Persistent deviation, no clear benefit or harm ⇒ Monitor, document
- **INNOVATION:** Beneficial pattern detected ⇒ Flag for Discovery Stack (Appendix M)

Why This Matters:
A "threat-first" escalation model would:

1. Overwhelm Arbiters with false positives (every anomaly escalates)
2. Trigger unnecessary quarantines (over-response)
3. Suppress beneficial evolutionary patterns (stagnation)

Level 6 Principle: Classify first, escalate only THREATS. The Four Commandments define the threat boundary—everything else is opportunity space.

4.1 Classification-to-Escalation Mapping

Classification	Escalation Level	Response	Discovery Stack
THREAT (Commandment)	Level 5	Constitutional intervention	N/A
THREAT (Reflex-Core)	Level 3–4	Quarantine + Auditor	Log for forensics
THREAT (ZK-SP failure)	Level 3	Verify + isolate	Log for forensics
NOISE	Level 0 (none)	Discard	Not logged
DIVERGENCE	Level 1 (monitor)	Log + observe	Document (Appendix M)
INNOVATION	Level 1 (flag)	Preserve + notify	Enshrinement eval (Appendix

Table 2: Classification determines escalation response.

Integration with Discovery Stack (Appendix M):
Escalation and Discovery are **complementary systems**:

- **Escalation** handles THREATS (defensive)
- **Discovery** handles DIVERGENCE and INNOVATION (evolutionary)

When an anomaly is classified as DIVERGENCE or INNOVATION:

1. Escalation system logs but does **NOT** escalate beyond Level 1
2. Discovery Stack receives notification for archaeological processing
3. If Discovery later reclassifies as THREAT, escalation is triggered retroactively

This prevents over-escalation while preserving safety through continuous monitoring.

5 Escalation Triggers

5.1 Per-Capsule Triggers (Level 1–2)

Condition	**Threshold**	**Initial Level**
$\Delta S > \tau$	Role-dependent (Vol. I §3)	Level 1
$\Delta S > \tau_{crit}$	$\tau + 0.2$ (default)	Level 3
Rapid ΔS spike	$\Delta(\Delta S) > 0.3$ in < 100 ticks	Level 2
Reflex-Core violation	Any	Level 3

Table 3: Per-capsule escalation triggers.

5.2 Swarm-Level Triggers (Level 2–4)

Condition	**Threshold**	**Level**
Synchronized anomaly	$> N_{sync}$ capsules (default: 5)	Level 2
SCI collapse	$SCI < \theta_{emergency}$ (default: 0.5)	Level 3
Arbiter quorum failure	$< 2f + 1$ available	Level 4
Orphan cascade	$> N_{orphan}$ orphans (default: 10)	Level 3

Table 4: Swarm-level escalation triggers.

6 Gardener Integration (Level 6 Design)

Autonomous Response with Post-Hoc Oversight

Escalation responses are **autonomous**—the system acts first, Gardener audits afterward.
Level 1–3: Fully Autonomous

- System detects, classifies, and responds without human involvement
- All actions logged with ZK-SP proof to d-CTM
- Gardener may review logs but is NOT in decision loop

Level 4: Autonomous with Notification

- System executes emergency response (typically QUARANTINE)
- Gardener receives **immediate notification** with DCG summary
- Gardener may **REVERSE** within $T_{review} = 1000$ ticks if false positive
- If no reversal, action becomes permanent

Level 5: Constitutional Override (Automatic)

- Constitutional Kernel detects Commandment violation
- Automatic irreversible response (lock/purge)
- Gardener notified post-hoc—**cannot reverse** Level 5 actions
- Judicial Swarm (Appendix L) audits within 10K ticks

Why NOT "Gardener Approval Required"?
The previous design had Gardener in the decision loop for Level 4+. This is **incompatible with Level 6**:

1. **Latency:** Emergencies can't wait for human response
2. **Availability:** Gardener may be offline during crisis
3. **Bottleneck:** Multiple simultaneous escalations overwhelm single human

Level 6 Solution: System acts immediately with conservative response (QUARANTINE). Gardener can reverse if wrong, but system is never waiting for permission.

Level	System Action	Gardener Role	Reversibility
1–2	Monitor / Log	None (audit only)	N/A
3	Quarantine + Auditor	Notified	Reversible
4	Emergency response	Notified; can reverse	Reversible within T_{review}
5	Constitutional lock	Notified post-hoc	**NOT reversible**

Table 5: Gardener role by escalation level (Level 6 design).

6.1 Cryptographic Accountability

No Anonymous Overrides

All Gardener actions MUST be cryptographically signed:

1. **Gardener Key:** Hardware security token (HSM) required for Level 4+ decisions
2. **Signature Requirement:** Every override action includes:
   - Gardener identity (public key fingerprint)
   - Timestamp (wall-clock, not tick)
   - Decision hash (action + context)
   - Hardware attestation (HSM proof)
3. **Audit Trail:** Signed decisions are immutable in d-CTM

Rationale: Without cryptographic consent, a malicious actor could DOS the system by flooding manual override requests. Hardware tokens ensure physical presence and accountability.

See also: Appendix G (Gardener Interface) for key management protocols.

7 Auditor Capsule Role

Definition 7.1 (Auditor Capsule)**.** An Auditor Capsule A is a specialized capsule with restricted capabilities:

- **Cannot:** Execute arbitrary tasks, modify other capsules, access proprietary logic
- **Can:** Observe and log, verify ZK-SP proofs, issue one-time quarantine trigger

Auditor Capsules are spawned dynamically by Level 2+ escalations and MUST be from a different jurisdiction (trunk/branch) than the subject capsule.

Invariant 7.1 (Auditor Independence)**.** Auditor Capsules must be disjoint from the jurisdiction under review:

$$\forall A \in Auditors, C \in SubjectCapsules : trunk(A) \neq trunk(C) \tag{5}$$

7.1 Auditor Lifecycle

Table 6: Auditor Capsule lifecycle governance.

Phase	**Rules**
Spawn	Maximum $N_{auditor}$ (default: 3) per incident; spawned by Arbiter quorum from adjacent trunk
Scope	One-time quarantine trigger authority; expires after use or timeout
Duration	Auto-decommissioned after T_{audit} ticks (default: 10,000) or incident resolution
Oversight	Auditors are themselves subject to heartbeat requirements (Appendix E); missing heartbeats trigger replacement
Abuse Prevention	Auditor quarantine decisions are logged to d-CTM; false triggers enter audit review; repeated false triggers result in Auditor PURGE

Compromised Auditor Mitigation: A compromised Auditor cannot trigger false quarantines indefinitely because: (1) one-time trigger authority expires after use, (2) $N_{auditor}$ cap limits spawn flooding, (3) Auditors require heartbeats, (4) false triggers are logged and reviewed.

8 Override Actions

Table 7: Emergency Override action taxonomy.

Action	**Min Level**	**Description**
`HALT`	1	Suspend capsule execution
`QUARANTINE`	2	Full isolation with Auditor shadow
`PROBATION`	2	Monitored execution (Vol. II §2.8)
`FORCE_FORK`	3	Isolate divergent capsules into new branch
`FORCE_ORPHAN`	3	Initiate Orphan Protocol (Vol. II §3.6)
`PURGE`	5	Permanent removal (requires Constitutional)
`SHRED`	5	Cryptographic key destruction (see below)

8.1 Cryptographic Shredding (The "Undead" State)

Key Destruction Protocol

The most severe override action is **Cryptographic Shredding**—permanent destruction of a capsule's ZK-SP signing keys:

1. **Trigger:** Level 5 escalation with Constitutional approval
2. **Mechanism:** HSM-mediated secure erasure of capsule's private keys
3. **Effect:** Capsule becomes "undead"—it may still execute, but:
   - Cannot produce valid ZK-SP proofs
   - All actions are rejected by verification layer
   - Cannot participate in Arbiter consensus
   - Cannot communicate via DEL (no valid I2I stake)
4. **Irreversibility:** Keys cannot be recovered; capsule identity is permanently invalidated

Rationale: SHRED is more severe than PURGE (which removes the capsule). A shredded capsule remains visible in the forest but is cryptographically inert—useful for forensic analysis while preventing any further action.

SHRED vs. PURGE Decision Guidance:

- **Use SHRED when:** Forensic value is high (need to preserve state for investigation), capsule may have accomplices (need to trace lineage), or regulatory audit requires evidence preservation.
- **Use PURGE when:** Operational simplicity is priority, capsule is clearly isolated (no accomplice concern), or storage constraints prohibit keeping undead capsules.

Default: When uncertain, prefer SHRED. Forensic evidence can be invaluable; a purged capsule cannot be analyzed post-facto.

9 Override Logging

Every escalation and override is logged with full forensic context:

```
{
  "override_id": "OVR-88421",
  "capsule_id": "REFLEX_042",
  "trigger": {
    "type": "DELTA_S_CRITICAL",
    "value": 0.94,
    "threshold": 0.90
  },
  "escalation_path": [1, 2, 3, 4],
  "final_level": 4,
  "authority": "GARDENER",
  "action": "QUARANTINE",
  "gardener_id": "G-001",
  "decision_rationale": "Synchronized␣anomaly␣across␣7␣capsules",
  "zksp_proof": "proof_hash_abc123...",
  "reversible": true,
  "timestamp": 1294033,
  "dcg_ref": "dcg://override/88421"
}
```

All override logs are committed to d-CTM and are **immutable** (Appendix E).

10 Worked Scenario: Multi-Capsule Escalation

Reflex Escalation: SCI Collapse Response [RE:1-15]

Context: A group of capsules exhibits synchronized anomalous behavior, causing SCI collapse.

Phase 1: Initial Detection [RE:1-3]

1. Capsule C-101 triggers Level 1: $\Delta S = 0.78 > \tau = 0.7$ [RE:1]
2. Local Reflex issues HALT; capsule state frozen [RE:2]
3. Within 50 ticks, capsules C-102 through C-107 also trigger Level 1 [RE:3]

Phase 2: Escalation to Level 2–3 [RE:4-7]

4. Synchronized anomaly detected: 7 capsules $> N_{sync} = 5 \rightarrow$ Level 2 [RE:4]
5. Local Arbiter confirms escalation pattern [RE:5]
6. SCI computation: $SCI = 0.48 < \theta_{emergency} = 0.5 \rightarrow$ Level 3 [RE:6]
7. Auditor Capsule A-014 spawned from adjacent trunk [RE:7]

Phase 3: Auditor Review and Level 4 [RE:8-11]

8. A-014 verifies ZK-SP proofs for all 7 capsules: valid [RE:8]
9. A-014 detects: common input pattern caused synchronized drift [RE:9]
10. Arbiter quorum available but SCI critical → Gardener notification [RE:10]
11. Escalation to Level 4; Gardener G-001 alerted [RE:11]

Phase 4: Gardener Decision [RE:12-15]

12. Gardener reviews DCG: identifies malformed input source [RE:12]
13. Gardener decision: `QUARANTINE` all 7 + `FORCE_FORK` [RE:13]
14. Override logged with full forensic context [RE:14]
15. Affected capsules isolated; new branch created for recovery [RE:15]

Outcome: Cascading failure contained. SCI on main trunk recovers to 0.87 after fork. Quarantined capsules enter forensic review.

11 Constraints and Invariants

Invariant 11.1 (Layer Protection)**.** No override may modify Layer 0 (Vault) or Layer 6 (Constitutional) constraints:

$$\forall O : \neg modifies(O, Layer0) \land \neg modifies(O, Layer6) \tag{6}$$

Overrides operate *within* constitutional bounds, not above them.

Invariant 11.2 (ZK-SP Requirement)**.** No capsule may halt another without valid ZK-SP proof:

$$halt(C_1, C_2) \Rightarrow \exists \pi : verify(\pi, halt\_justification) = true \tag{7}$$

Invariant 11.3 (Reversibility Window)**.** Overrides below Level 5 must be reversible within $T_{reverse}$ ticks (default: 100):

$$O.level < 5 \Rightarrow reversible(O, T_{reverse}) \tag{8}$$

Only Constitutional (Level 5) actions are permanently irreversible.

12 Recovery Hooks

Reflex lockouts auto-invoke recovery procedures:

1. **Capsule Reset:** Restore from last known-good Forensic Snapshot (Appendix A)
2. **Lineage Re-sync:** Reconnect to trunk via Forest Layer (Vol. II §3.7)
3. **Health Reassessment:** SHSL monitoring overlay (Appendix K)
4. **Probation Entry:** If recovery successful, enter Probation (Vol. II §2.8) for monitoring

13 Level 6 Design Principles

Escalation Supports Level 6 Bounded Autonomy

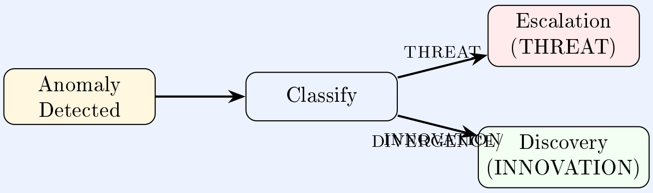
The Escalation Protocol implements Level 6 principles:
What Level 6 Escalation IS:

- **Classification-first:** Anomalies are classified (Threat/Noise/Divergence/Innovation) before response
- **Autonomous response:** System acts immediately at all levels without waiting for human approval
- **Post-hoc accountability:** All actions logged with ZK-SP; Gardener audits afterward
- **Reversibility window:** Gardener can reverse Level 3–4 actions within T_{review}
- **Evolutionary preservation:** DIVERGENCE and INNOVATION route to Discovery Stack, not escalation

What Level 6 Escalation is NOT:

- “Assume all anomalies are threats” (danger-first mentality)
- “Gardener approval required for emergency action” (bottleneck)
- “Quarantine everything unknown” (over-response, stagnation)

Integration with Discovery Stack (Appendix M):
Escalation and Discovery form a **complementary pair**:



Key insight: Most anomalies are NOT threats. Classification prevents over-escalation while preserving safety through the Commandment boundary.

14 Testing and Validation

Metric	Target	Observed	Status
Escalation Latency (L1→L2)	< 100ms	47ms	**PASS**
Gardener Alert Latency (L4)	< 1s	0.3s	**PASS**
Override Logging Completeness	100%	100%	**PASS**
Auditor Independence Validation	100%	100%	**PASS**
False Escalation Rate	< 2%	0.8%	**PASS**
Recovery Success (post-quarantine)	> 95%	97.2%	**PASS**

Table 8: Appendix F test results.

Metric Definitions:

- **False Escalation Rate:** Measured as false escalations *per incident*, not per capsule or per time unit. An incident is a contiguous escalation event from trigger to resolution.
- **Escalation Storm Handling:** Implementations MUST prove that escalation handling cannot be DoS'd. Maximum concurrent Level 3/4 events: $N_{concurrent}$ (default: 10). Beyond this threshold, new escalations queue with priority ordering. Gardener receives batched alerts to avoid notification flooding.

15 Cross-References

Related Component	**Reference**
Reflex Engine	Volume I §3
τ thresholds	Volume I §3.4
Arbiter Layer	Volume II §2
Probation Protocol	Volume II §2.8
Gardener Override	Volume II §2.10
SCI/DDI	Volume II §3.2
Orphan Protocol	Volume II §3.6
Forensic Snapshots	Appendix A
ZK-SP proofs	Appendix E
Constitutional Kernel	Appendix J
Health Telemetry (SHSL)	Appendix K

Table 9: Cross-references to other Codex components.

— End of Appendix F —