Appendices

These appendices supply the reproducible scaffolding for the study: the pre-registered **Analytical Question Battery** that defines the inquiry space; concise **Deductive Memo** abstracts that constitute the text-as-data corpus; a **Codebook summary** with operational definitions and inclusion/exclusion rules to secure construct clarity; a **Coded Matrix** synopsis (salience, co-occurrence, exemplars, reliability) that renders the qualitative signal auditable; and a compressed **Design Pattern Catalog** that translates high-salience mechanisms into implementable artefacts. Together they fix the unit of analysis, preserve versioned provenance from principle to policy, and provide a direct crosswalk to Q2 simulation so that claims can be replicated, compared across settings, and subjected to severe tests.

Appendix A — Analytical Question Battery (20 Items)

1. **Requisite Variety:** Do the agent + escalation chain collectively match environment demand variety?
2. **Attenuation Mix:** Is intake attenuation calibrated to minimize (misses + rework + delay) subject to risk bounds?
3. **S2 Coordination:** Do parallel agents operate under an explicit S2 protocol to prevent conflicts/duplication?
4. **S2 vs S3 Separation:** Are coordination (S2) and assurance/control (S3) roles cleanly separated?
5. **S4 vs S5 Separation:** Is intelligence/experimentation (S4) distinct from policy/identity (S5)?
6. **Gain Tuning:** Are controller gains tuned (with anti-windup/rate limits) for the process?
7. **Feedforward Gating:** Are anticipatory actions gated by forecast quality and paired with rollback?

8. **Autonomy Bounds:** Are agent autonomy envelopes explicit, machine-enforced, and recursion-aligned?
9. **Escalation Thresholds:** Are magnitude×duration×novelty thresholds quantified and honored, with algedonic bypass?
10. **Observability:** Is state reconstructible with timely, complete telemetry?
11. **World Model:** Do regulators embody a domain model adequate for their control tasks?
12. **Variety-Aware Routing:** Is case assignment capability-matched rather than FIFO/round-robin under heterogeneity?
13. **Buffers:** Are buffers sized to variance/tails (not means) with overload throttles?
14. **Degeneracy:** Are heterogeneous backups deployed for correlated failure resilience?
15. **Monitor vs Operator:** Are monitoring and intervention roles separated to reduce operator-induced incidents?
16. **Constraint-First:** Are constraints encoded as first-class setpoints and guards?
17. **Experimentation Cadence:** Is there a fixed cadence with control groups and auto-rollback?
18. **Objectives vs Policies:** Are ends (objectives) separated from means/constraints (policies) to reduce reward hacking?
19. **S3 Assurance Coverage:** Does S3 cover the full compliance variety faced by the organization?
20. **Steward Capacity:** Is human stewardship bandwidth/skill matched to agent autonomy?

Appendix B — Deductive Memos (Abstracts, 2–3 sentences each)

Q1 Requisite Variety. Matching environmental variety via attenuation/amplification reduces breaches and escalations; under-variety

leaks as exceptions. Counter-case: forgiving domains with slack. Evidence to overturn: equalized variety fails to improve outcomes.

Q2 Attenuation Optimum. Intake filtering exhibits a U-shaped cost over (misses + rework + delay). Over-filtering hides hazards; under-filtering overloads. Overturn if costs fall monotonically across broad threshold sweeps.

Q3 S2 Coordination. S2 dampens peer interference; without it, conflicts and duplicate work rise. Overhead can dominate in embarrassingly parallel work. Overturn if removing S2 leaves conflict metrics unchanged.

Q4 S2 vs S3 Separation. Coordination (S2) ≠ assurance/control (S3); fusion causes micromanagement or gaps. Overturn if fused role yields fewer rollbacks/defects at equal risk.

Q5 S4 vs S5 Separation. Exploration (S4) must be institutionally distinct from policy (S5) to avoid capture/freeze. Overturn if integrated S4/5 improves viability without policy breaches.

Q6 Gain Tuning. Untuned controllers with deadtime oscillate; PID-like tuning with limits stabilizes KPIs. Overturn if naïve high-gain achieves equal/better tails without incidents.

Q7 Feedforward Gating. Forecast-gated feedforward beats feedback-only during predictable surges; without gates it adds bias. Overturn if credible forecasts provide no peak improvement.

Q8 Autonomy Bounds. Machine-enforced envelopes reduce off-policy actions and unsafe generalization. Overturn if relaxed/implicit bounds reduce incidents at equal risk.

Q9 Escalation Thresholds. Quantified thresholds beat discretion; algedonic path for irreversibles. Overturn if thresholds raise severe incidents vs expert discretion.

Q10 Observability. Rich telemetry + provenance shortens MTTR and improves tuning/accountability. Overturn if upgrades leave MTTR/recurrence flat.

Q11 World Models. Model-based regulators outperform rule-only in complex regimes. Overturn if rule-only matches in high-variety settings.

Q12 Variety-Aware Routing. Capability-matched routing lifts FCR; fairness must be monitored. Overturn if FIFO matches performance under heterogeneity.

Q13 Buffers. Variance-sized buffers tame P95/99 latency spikes; means mislead. Overturn if JIT achieves equal tail control in heavy-tail arrivals.

Q14 Degeneracy. Heterogeneous backups reduce correlated outage risk. Overturn if homogeneous redundancy equals performance under correlated faults.

Q15 Monitor vs Operator. Separating observing from actuating reduces operator-induced incidents. Overturn if merged roles lower incidents/MTTR without error spikes.

Q16 Constraint-First. Hard guards and setpoint bands reduce tail loss with acceptable throughput penalty. Overturn if severe incidents don't fall post-hardening.

Q17 Experimentation Cadence. Release trains with rollback prevent drift/regressions. Overturn if ad-hoc changes outperform with equal safeguards.

Q18 Objectives vs Policies. Separating ends from means reduces reward hacking. Overturn if unified optimization meets safety without violations.

Q19 S3 Coverage. Assurance must mirror full compliance variety to avoid repeat findings. Overturn if expanded coverage fails to reduce repeats/audit time.

Q20 Steward Capacity. Oversight must match autonomy; otherwise bottlenecks or rubber-stamps appear. Overturn if reduced stewardship has no measurable downside.

Appendix C — Codebook Summary (Themes → Codes, one-line definitions)

Theme A: Variety Management

- **RV_MATCH:** Matching regulator capability to demand variety.
- **ATTENUATION:** Intake filtering/structuring to reduce incoming variety.
- **AMPLIFICATION:** Expanding capability via tools/permissions/data.
- **VAR_ROUTING:** Capability-matched case assignment.
- **BUFFERING:** Slack sized to variance/tails.
- **DEGENERACY:** Heterogeneous backups for correlated risks.

Theme B: Control Architecture & Dynamics

- **FEEDBACK_LOOP:** Reactive measurement→action control.
- **FEEDFORWARD:** Anticipatory action from forecasts.
- **GAIN_TUNING:** Parameter tuning incl. anti-windup/rate limits.
- **DELAY_COMPENSATION:** Handling deadtime/latency.
- **OSCILLATION:** Limit cycles/overshoot symptoms.
- **SETPOINT_BANDS:** Homeostatic bands/guardrails.

Theme C: VSM Structure & Governance

- **S2_COORDINATION:** Peer synchronization/conflict damping.
- **S3_ASSURANCE:** Allocation, rollback, compliance checks.
- **S4_INTEL:** Forecasting/experimentation/horizon scanning.
- **S5_POLICY:** Identity, objectives, constraint selection.
- **RECURSION_ALIGNMENT:** Cross-level goal/policy coherence.
- **AUTONOMY_BOUNDS:** Explicit authority/constraint envelope.
- **ESCALATION_THRESHOLDS:** Quantified handoff triggers.
- **MONITOR_VS_OPERATOR:** Separation of observing vs acting.

Theme D: Observability, Modeling & Drift

- **OBSERVABILITY:** Telemetry sufficiency and freshness.

- **PROVENANCE:** Decision trails for auditability.
- **WORLD_MODEL:** Embodied domain model for control.
- **DRIFT_MONITORING:** Detect/correct model/goal/process drift.
- **FORECAST_QUALITY:** Fitness of predictions for control.

Theme E: Constraints, Safety, Compliance & Ethics

- **CONSTRAINT_FIRST:** Constraints as first-class setpoints/guards.
- **COMPLIANCE_VARIETY:** Assurance coverage breadth.
- **EXPLAINABILITY:** Stakeholder-appropriate reasons.
- **FAIRNESS_BOUNDS:** Declared fairness constraints and monitors.
- **ALG_HARD_STOP:** Algedonic hard-stop for irreversibles.

Theme F: Experimentation, Change & Incentives

- **EXPERIMENTATION_CADENCE:** Fixed, governed release rhythm.
- **ROLLBACK_PROTOCOL:** Pre-declared reversion criteria.
- **REWARD_HACKING:** Proxy-gap and off-policy gaming.

Appendix D — Coded Matrix Summary (Statistics + Representative Rows)

D1. Statistics

- **Segments coded:** 60 (3 per memo: baseline, counter-argument, implication).
- **Mean codes/segment:** 1.7 (multi-coding permitted, max 3).
- **Top codes:** S2_COORDINATION (6), S3_ASSURANCE (6), S5_POLICY (6), ATTENUATION (5), AUTONOMY_BOUNDS (5), CONSTRAINT_FIRST (5).
- **Top co-occurrences:** ATTENUATION×AMPLIFICATION; S2_COORDINATION×S3_ASSURANCE; S4_INTEL×S5_POLICY; DELAY_COMPENSATION×GAIN_TUNING.

D2. Representative Rows (verbatim excerpts)

- **Q1-1 (Baseline):** "By Ashby, a regulator (agent + escalation chain) must match environmental variety; otherwise disturbances leak as escalations…" → **RV_MATCH**.
- **Q2-1 (Baseline):** "Total cost is typically U-shaped: too little attenuation overloads regulators; too much hides rare but critical signals…" → **ATTENUATION**.
- **Q4-1 (Baseline):** "S2 harmonizes; S3 allocates resources and assures compliance." → **S2_COORDINATION; S3_ASSURANCE**.
- **Q6-1 (Baseline):** "With delays and noise, untuned controllers overshoot and oscillate." → **DELAY_COMPENSATION; GAIN_TUNING; OSCILLATION**.
- **Q10-3 (Implication):** "Define a minimum viable observability spec… treat provenance coverage as a first-class KPI." → **OBSERVABILITY; PROVENANCE**.

D3. Reliability (pilot)

- **Krippendorff's α (nominal):** 0.81 on 10% double-coded sample.
- **Hot spots:** S2 vs S3 vs S5 boundaries; resolved via tie-break heuristics and exemplars.

Appendix E — Design Pattern Catalog (Compressed Summaries)

1. **Variety-Gated Intake** — Calibrate intake schemas/thresholds to minimize (misses + rework + delay) under risk bounds. **Metrics:** FN rate, rework %, P95 intake→first-action. **Link to Q2:** attenuation strength, subtype prevalence, miss/rework curves.
2. **Capability-Aware Routing** — Assign by capability/confidence with fairness monitors. **Metrics:** FCR, P90 handle time (complex), parity gaps. **Q2:** routing policy, signal accuracy, heterogeneity.

3. **Autonomy Envelope with Machine Checks** — Pre-/post-act guards and novelty detectors within explicit authority. **Metrics:** off-policy incident rate, autonomy-breach alerts, rollback frequency. **Q2:** bound strictness, novelty rate, discretion level.

4. **Thresholded Escalation + Algedonic Channel** — Magnitude×duration×novelty thresholds with hard-stop for irreversibles. **Metrics:** time-to-escalation, overrides, tail-loss percentiles. **Q2:** threshold levels, shock injections, algedonic enable.

5. **S2 Synchronization Protocol** — Locks, back-pressure, conflict rules for peers. **Metrics:** conflict incidents/1k actions, duplicate-action rate, WIP CV. **Q2:** coupling, S2 on/off, update cadence.

6. **Constraint-First Governor** — Encode constraints as setpoints/bands; drill kill-switches. **Metrics:** severe-incident rate, constraint-violation count, throughput delta. **Q2:** constraint hardness, penalty weights, algedonic readiness.

7. **Forecast-Gated Feedforward** — Enable anticipatory actions only when forecast gates pass; auto-rollback otherwise. **Metrics:** peak P95 latency, surge breach rate, forecast error vs benefit. **Q2:** MAPE/coverage gates, buffers, enable flag.

8. **Tail-Risk Buffering** — Size buffers to variance/tails; add throttles and graceful degradation. **Metrics:** P95/P99 latency, overload frequency, buffer utilization. **Q2:** buffer size, arrival variance/tailness.

9. **Mixed-Mechanism Degeneracy** — Heterogeneous backups with arbiter logic for correlated risks. **Metrics:** joint outage probability, failover time, performance under faults. **Q2:** heterogeneity, correlation, arbiter rule.

10. **Experimentation Cadence + Auto-Rollback** — Fixed release trains with pre-registered guardrails and automated reversions. **Metrics:** regressions/change, rollback rate, change MTBF. **Q2:** cadence, guardrail hardness, change volume.

11. **Gain-Tuned & Delay-Compensated Control** — PID tuning, derivative clamps, slew-rate limits, setpoint bands. **Metrics:** overshoot/undershoot counts, settling time, control-effort variance. **Q2:** gains, deadtime, noise.

12. **Observability & Decision-Provenance Spine** — Minimum viable events; freshness SLAs; replayability. **Metrics:** provenance coverage, MTTR, repeat-incident rate. **Q2:** telemetry richness, log latency, sampling rate.

Usage & Versioning

- **Traceability:** Each appendix item is versioned (v1.0) and should be frozen before Q2 tests; subsequent edits require a change log.
- **Cross-walk:** All patterns list metrics and Q2 knobs to enable severe tests; acceptance thresholds should be pre-registered to avoid post-hoc rationalization.