**The Trident G-Loop: A Unified Cognitive Architecture for Adaptive, Near-Critical Intelligence**

# **Abstract**

* One loop, eight phases, operating inside a **Ψ-band** (near-critical regime) by regulating two interacting axes: a **Cognitive Resilience axis** (difficulty set-point **F\*** and precision-weighted gap **ΔF̂ₜ = κ·(ℰₜ − F\*)**) and an **Inference axis** (precision-bias **b = β\_d(d − d\*)** and **temperature T(χ)**).
* **Trident geometry**: subcritical **Autopilot shaft** → near-critical **branch point** (ℰ≈F\*) → **Control** (compress/exploit) vs **Creative** (decompress/explore) prongs, coordinated by **Salience** (monitor/orient/reset).
* Two-sentence coda on the **Ψ-dynamic range of 𝒢** (𝒢 = r × G\_f): η-driven **F\*** ratcheting and controllable **b/T/λ** spread; list telemetry: **Ω\_Ψ** (band occupancy), **τ½** (recovery half-life), **H\_Ψ** (hysteresis), **μ\_{F\*}** (set-point mobility).
* Reasoning in Trident G is realised by two cooperating WM workspaces: a relational, map-based LT-WM (FPCN-A↔hippocampus/DMN) and a rule/sequence ST-WM (FPCN-B↔DAN), with salience arbitrating the hand-off inside the Ψ-band.

# **1. Introduction**

1.1 The challenge of flexible intelligence (stability ↔ flexibility).  
 1.2 The **Ψ-band principle**: benefits of operating between order and chaos; band (not point) via hierarchical/heterogeneous networks.  
 1.3 The **G-Loop & Trident metaphor** (Autopilot shaft, branch, two prongs, salience hilt).  
 1.4 The **Two Axes**: **Resilience** {F\*, ΔF̂ₜ = κ(ℰₜ − F\*)} and **Inference** {b = β\_d(d−d\*), T(χ), λ}; preview **η** (competence/recovery) and **χ** (inconsistency/volatility).

**1.5 Evolutionary rationale.** We view human reasoning as an extension of Phase-3 mapping and Phases-4/5 execution: **abduction + counterfactuals** enable safe search in uncertain niches, while **deduction + induction** compress and transmit validated structure. Salience arbitrates these modes to keep operation near **Ψ**, yielding adaptive intelligence in variable environments.  
 1.6 Roadmap.

# **2. Theoretical Foundations & Core Commitments**

2.1 **Policy objectives**: neutral score **𝒥(a|s)**; Active Inference uses **𝒥 = −G\_{\mathrm{EFE}}**, classical decision uses **𝒥 = \mathrm{EV}(a|s)**; selection via **π(a|s) ∝ exp(𝒥/T(χ))**; precision control via **b** and **T(χ)**.

* Use **CVaR\_q** (e.g., q=0.95) or a **τ-quantile** value function when tail-shape telemetry (Sec. 6) flags heavy-tail/volatility regimes.
* Tie the switch explicitly to telemetry thresholds (ζ/J-ratio/ES from Sec. 6) so readers see the closed loop: “The risk-sensitive objective engages when ζ<ζ\* or J-ratio>J\* or ES\_q>ES\*

2.2 **Relational core (φ)**: hippocampal–prefrontal **predictive map** (SR-style), value-landscaped by **EV** or **−G\_{\mathrm{EFE}}**; learned translators to modalities.  
 2.3 **Metacognition as cross-loop consistency**: **χ\_{meta}** monitors inter-loop inconsistency; brief monitor pulses to re-centre.  
 2.4 **Multi-timescale architecture**: fast/medium/slow controllers; rhythms (γ/θ/δ) as dynamic weights (non-exclusive).

# **3. Controllers & State Variables (the meta-cognitive toolkit)**

3.1 **State signals**: **η** (competence/recovery), **χ** (uncertainty/inconsistency).  
 3.2 **Controllers**:  
 • **F\*** thermostat (Resilience axis; branch set-point).  
 • **d, d\*** (dopamine tilt & neutral set-point) → **b = β\_d(d − d\*)** (Inference tilt).  
 • **T(χ)** (policy temperature), **λ, λ\*** (generalisation width & target), **α** (plasticity), **κ** (gate gain in ΔF̂), **σ** (reasoning budget), **ν** (decompression/novelty budget).  
 3.3 **η → F\*** coupling: competence-driven upward drift (auto-difficulty).

# **4. Core Hypotheses (with short tests)**

**H1 — SOC-for-Inference.** Homeostatic plasticity drives hierarchical networks into a **Ψ-band** that maximises information capacity/dynamic range required for good policy performance; loops track **F\*** and then choose policies by **𝒥** (EV or −G\_{\mathrm{EFE}}). *Test:* shift E:I toward/away from balance and track Ψ-metrics + performance/𝒥-proxies.  
 **H2 — Trident Branching.** Near **ℰ≈F\*** and elevated **χ**, **ΔF̂** and salience arbitrate a bifurcation into **Control** (compress) vs **Creative** (decompress), with monitor pulses to re-centre **b**/**T**. *Test:* SN-led switches into FPCN-B (Control) vs FPCN-A/DMN (Creative) with prong-specific telemetry.  
 **H3 — Global + Per-loop Ψ-bands.** Multiple G-loops operate with loop-specific **Ψ\_k** windows whose intersection defines a global **Ψ**; **χ\_{meta}** triggers monitoring when loops disagree. *Test:* show Griffiths-like banding across scales and salience-preceded re-entry.  
 **H4 — Ψ-Dynamic-Range of 𝒢.** **𝒢 = r × G\_f** reflects (i) **Resilience range** (tolerable |ℰ−F\*| and η-driven **μ\_{F\*}**) and (ii) **Inference range** (controllable **b/T/λ** span with clean re-entry), aggregated across loops. *Test:* derive a Ψ-profile (Ω\_Ψ, τ½, H\_Ψ, μ\_{F\*}) and relate to Gf/Gc; show training gains.  
 **H5 — Two Creative Routes.** Exploration enters via (i) boredom/under-challenge (soft D2) and (ii) error/volatility (hard D2), with distinct arousal/ACC signatures. *Test:* dissociate behavioural/physiological markers for the two routes.  
 **H6 — Dual WM Workspaces.** FPCN-A↔hippocampus supports long-horizon relational “map” (Creative/Plan), FPCN-B↔DAN supports rule/sequence execution (Control), with salience arbitration. *Test:* mode-specific coupling shifts across phases.

# **5. The Eight Phases of the G-Loop (navigating the Trident)**

1. **Sense & Move** (estimate **ℰₜ**).
2. **Gap-check & Mode Branch** (compute **ΔF̂ₜ = κ·(ℰₜ − F\*)**; arbitrate via salience using **χ, χ\_{meta}**).
3. **(De)Compress φ** (Control: compress → **η↑**; Creative: decompress under **λ↓/ν↑**).
4. **Decide/Gate** (softmax over **𝒥** with **𝒥 = EV** or **−G\_{\mathrm{EFE}}**; **T(χ)** sets exploration; **F\*** does not enter softmax).
5. **Execute** (policies expected to reduce **χ** and align ℰ→F\*).
6. **Reflect/Update** (adjust **F\***, **b**, **T**; update **η, χ**).
7. **Consolidate (λ)** (propagate validated structure; quarantine speculative bits).
8. **Reset** (restore Ψ-band rhythms; prep next cycle).

# **6. Telemetry, Predictions & Falsification**

* **Band signatures:** avalanche scaling/dynamic range; metastability peaks near branch; rare-region effects (Griffiths-like).
* **Prong states:** SN-led switches → FPCN-B (Control) vs FPCN-A/DMN (Creative).
* **Two exploration routes:** boredom vs error/volatility with distinct arousal/ACC patterns.
* **Map (de)compression evidence:** SR-like updates; pattern separation/completion.
* **Ψ-profile metrics:** **r**, **G\_f**, **Ω\_Ψ**, **τ½**, **H\_Ψ**, **μ\_{F\*}**. *Falsify where moving toward near-criticality fails to improve Ψ-metrics or where SN activity does not precede re-entry.*

**Tail-shape telemetry.**

**ζ (tail index):** Hill α\_Hill for power-law–like tails (optionally report excess kurtosis κ\_ex). **Lower α\_Hill ⇒ heavier tail.**

**J-ratio (one-big-jump index):** J=max⁡\_t∣ΔF\_t∣/∑\_t∣ΔF\_t∣J=\max\\_t|\Delta F\\_t| \Big/ \sum\\_t |\Delta F\\_t|J=max\_t∣ΔF\_t∣/∑\_t∣ΔF\_t∣ over the window; **higher J ⇒ single-shock dominance**.

**ES\_q(ΔF):** Expected Shortfall of ΔF at level q (e.g., q=0.95).

**Control law:** If α\_Hill<ζ\*, or J>J\*, or ES\_q>ES\*, then **raise T(χ)** (broader sampling), **lower λ** (sandbox Creative), trigger a **Monitoring pulse**, and recentre **b**; relax when metrics fall below thresholds.

**Prediction:** Tail alarms should precede **T(χ)↑, λ↓, Monitoring↑** and reduce **ES\_q(ΔF)** and band exits.

# **7. Minimal Computational Program (plus ablations)**

**Hybrid instantiation.** Wilson–Cowan E/I nodes (local Hopf-edge control) with Kuramoto-style phase coupling on a hierarchical-modular / empirical connectome.

**Controller mapping.** F∗F^\*F∗: slow homeostasis to target partial synchrony; bbb: subnetwork gain/tilt (via d−d∗d-d^\*d−d∗); T(χ)T(\chi)T(χ): input-noise / frequency-jitter gate; λ/α\lambda/\alphaλ/α: propagation & plasticity knobs; J\mathcal{J}J: EV/−G\_{EFE} or **risk-sensitive (CVaR/quantile) when tail flags fire**.

**Readouts (additions in bold).**

* Global/meso order parameter R(t)R(t)R(t); power-law exponents for (de)synchronisation; E:I telemetry; recovery E ⁣→ ⁣F∗\mathcal{E}\!\to\!F^\*E→F∗ and τ1/2\tau\_{1/2}τ1/2​ after shocks.
* **Tail-shape metrics on performance residuals ΔF\Delta FΔF:** **ζ\zetaζ** ===(**αHill\alpha\_{\text{Hill}}αHill​**, **κex\kappa\_{\text{ex}}κex​**); **J-ratio** J=max⁡t∣ΔFt∣/∑t∣ΔFt∣J=\max\_t|\Delta F\_t|/\sum\_t|\Delta F\_t|J=maxt​∣ΔFt​∣/∑t​∣ΔFt​∣; **ES\_q(ΔF\Delta FΔF)** (e.g., q=0.95q=0.95q=0.95).
* Gate & safety telemetry: T(χ)T(\chi)T(χ), λ\lambdaλ, bbb, **Ψ-band exits** (count/length).

**Protocols & stress-tests.**

1. **Baseline Gaussian (light-tail)**: finite-mgf noise; expect CLT regime—risk-neutral J\mathcal{J}J acceptable near mean.
2. **Heavy-tail / “one-big-jump”**: Pareto-like or α-stable noise + rare shocks; expect **risk-sensitive J\mathcal{J}J** to reduce ES\_q, band exits, and τ1/2\tau\_{1/2}τ1/2​.
3. **Injection tests**: single large outlier vs many small; verify **J-ratio↑ ⇒ T(χ)↑,λ↓T(\chi)↑, \lambda↓T(χ)↑,λ↓** and faster safe re-entry.  
    **Prediction:** Only the **tail-aware** agent maintains high **Ψ-occupancy**, lower **ES\_q(ΔF\Delta FΔF)**, fewer/shorter Ψ exits, and shorter τ1/2\tau\_{1/2}τ1/2​ under heavy tails.

**Ablations.** (i) **No F∗F^\*F∗** thermostat → over/under-challenge; unstable R(t)R(t)R(t); slow E ⁣→ ⁣F∗\mathcal{E}\!\to\!F^\*E→F∗.  
 (ii) **No χ ⁣→ ⁣T\chi\!\to\!Tχ→T** gate → brittle exploration/exploitation; variance spikes; excess Ψ exits.  
 (iii) **No η\etaη** gate (monitoring pulse) → poor regime re-entry; longer τ1/2\tau\_{1/2}τ1/2​.  
 **(iv) No risk-sensitive J\mathcal{J}J** (force EV/−G\_{EFE}) → **catastrophic failure modes in heavy-tailed environments**: **ES\_q(ΔF\Delta FΔF)↑**, J-ratio events propagate, Ψ-occupancy falls, despite intact F∗F^\*F∗ and χ ⁣→ ⁣T\chi\!\to\!Tχ→T.

**Falsification criteria.** If enabling tail-aware J\mathcal{J}J and tail telemetry **does not** (vs ablation iv) lower ES\_q, shorten τ1/2\tau\_{1/2}τ1/2​, and reduce Ψ exits in the heavy-tail condition—claim is weakened.

# **8. Reasoning, Working Memory & Constraints**

* **Deduction/Induction** = Control; **Abduction** = Creative; **Counterfactuals** at Decide/Act; **Analogy** bridges Creative→Control.
* **Constraint satisfaction:** feasibility gating/propagation (hard/soft/chance constraints) to guard Creative proposals and tighten Control propagation.
* **Dual WM:** long-term WM (FPCN-A↔hippocampus/DMN) vs short-term WM (FPCN-B↔DAN); salience hand-off. (Note - Both LT-WM (map) and ST-WM (rule/sequence) can operate in **either** creative (D2-biased) or control (D1-biased) mode; **D1:D2 sets processing style, not locus**—salience arbitrates the hand-off.)
* On this view, **reasoning is the control law** for staying inside the Ψ-band: Creative (abduction/analogy + counterfactuals) expands the hypothesis/policy space when χ is high; Control (deduction/induction + constraints) compresses and propagates validated structure for skillful action.”

# **9. Neuromodulatory & Rhythmic Implementation**

* **DA (D1/D2)** → **d, d\*** → **b** (stability–flexibility tilt); **LC–NE** → **T(χ)** (exploit–explore gain); **ACh** → precision/learning; **α/θ/γ** rhythms as dynamic weights.

# **10. Relation to Prior Work**

* Active Inference, predictive coding, meta-RL, MDN/FPN accounts—**what’s shared vs what’s new**: the **F\*** thermostat, **η/χ** gating, the **Ψ-band + Trident branching** geometry, and concrete, falsifiable telemetry.

# **11. Applications & Deployment Contract**

* Learning/training, digital therapeutics, decision support; include a deployment table aligned to controllers (**χ→T**, **F\***, **η-gate**, **b**, **λ**) and standard governance artefacts (model cards, risk registers, etc.).

# **12. Limitations & Ethics**

* What’s solid vs **partly supported/speculative** (e.g., equating dynamical criticality with an EFE set-point; global neutral D1:D2; χ/η as cross-loop constructs). State tests and caveats explicitly; clinical mappings as computational phenotypes, not diagnoses.
* **Supported:** near-critical banding / Griffiths-like effects; E:I tuning; salience-driven switching; Gf↔MD/FPN, Gc↔DMN tendencies; metastability–creativity links; D1/D2 stability–flexibility trade-off; LC/NE as temperature-like.
* **Promising/speculative:** mapping dynamical criticality to an EFE-compatible set-regime; a global neutral D1:D2 (d\*); χ,η as cross-loop constructs; dual exploration routes within one loop.

# **13. Conclusion**

* Re-state the **Trident** synthesis: Autopilot shaft → near-critical branch → prongs + salience hilt; two axes negotiating inside a **Ψ-band** to yield adaptive intelligence; concrete predictions and a compact simulation path.

# **Appendices**

* **A. Mathematical formulae** (F\* thermostat, **b** updates via **d−d\***, choice policy with **𝒥**, **λ** controller; normal-form cartoon for branching).
* **B. Glossary of variables** (η, χ, F\*, d/d\*, **b**, T, λ, α, κ, σ, ν) and network labels (FPCN-A/B, DMN, DAN, SN).