BB-BPRF — Formal Mathematical Model (v1.1, refreshed)

Purpose. A self-contained formalization of the Bias-Blind Behavioral Pattern Recognition Framework (BB-BPRF): objects, operations, policy/typing discipline, proof obligations, and complexity bounds. Designed for filing appendices, examiner briefings, and peer review.

---

0. Table of Symbols

- $\mathcal{S}$ — set of subjects; $s \in \mathcal{S}$
- $\mathcal{M}$ — set of modalities; $m \in \mathcal{M}$
- $X_{s,m}(t) \in \mathbb{R}^{d_m}$ — feature vector for subject s, modality m, time t
- $X_s(t) = \bigoplus_{m \in \mathcal{M}_s} X_{s,m}(t) \in \mathbb{R}^{d_s}$ — concatenated features
- $\mu_s, \sigma_s \in \mathbb{R}^{d_s}$ — per-subject robust location/scale
- $C_s \in [0,1]^k$ — cultural/context vector (k small, e.g., 4)
- $E \equiv (\mu^E, \sigma^E, C^E)$ — Ephemeral Initialization Vector (EIV) priors
- $w_{EIV}(t) = \max(0, 1 - t/T)$, $w_{ind}(t) = 1 - w_{EIV}(t)$ — decay weights
- $A_s \in \mathbb{R}^{d_s \times d_s}$ (diagonal), $D_s, T_s \in \mathbb{R}^{d_s}$, $R_s \in \mathbb{R}^{d_s \times d_s}$ (sparse)
- $X_s^{adj}(t) = A_s(D_s \circ X_s(t)) + T_s + R_s X_s(t)$
- $\mathcal{V} = \mathcal{V}_{ind} \uplus \mathcal{V}_{proc} \uplus \mathcal{V}_{forbid}$ — addressable variables by class
- $\mathscr{L}$ — library of subject-local operators
- $G_s(t)$ — computation DAG for subject s at time t
- P — policy graph over $\mathcal{V}$ with labels {ind, proc, forbid}

---

1. Formal Objects

1.1 Signals and State

Let time $t \in \mathbb{N}$. For each $s \in \mathcal{S}$, modality m has dimension d_m and observation $X_{s,m}(t) \in \mathbb{R}^{d_m}$. The active feature is $X_s(t) = \bigoplus_{m \in \mathcal{M}_s} X_{s,m}(t) \in \mathbb{R}^{d_s}$. Maintain state $(\mu_s(t), \sigma_s(t), C_s(t))$. Dispersion is strictly positive component-wise ($\sigma_s > 0$).

1.2 EIV-weighted Baseline

For horizon $T>0$, [$\mu_s^{eff}(t) = w_{EIV}(t)\, \mu^E + w_{ind}(t)\, \mu_s(t)$, $\sigma_s^{eff}(t) = w_{EIV}(t)\, \sigma^E + w_{ind}(t)\, \sigma_s(t)$.] EIV influence vanishes after $t \geq T$ ($w_{EIV}(t)=0$).

1.3 Policy Graph and Typing

Let $P=(N,E_P,\lambda)$ where $N=\mathcal{V}$ and $\lambda: \mathcal{V} \to \{ind, proc, forbid\}$. A program expression e is *well-typed* iff: 1) (Single-Subject Rule, SSR) e references variables for at most one s. 2) e references no variable v with $\lambda(v)$=forbid. 3) Every operator $f \in \mathscr{L}$ used by e is *subject-local* (all arguments pertain to the same s) and total on its domain.

1.4 Closures

For subject s and seed $U \subseteq \mathcal{V}\_{ind}(s)$, define the population-of-one closure $\mathcal{C}\_s(U)$ as the least set s.t. $U \subseteq \mathcal{C}\_s(U)$ and if $f \in \mathscr{L}$ is subject-local with args $\subseteq \mathcal{C}\_s(U)$, then $f(args) \in \mathcal{C}\_s(U)$. If any arg $\in \mathcal{V}\_{forbid}$ or references $s' \neq s$, $\mathcal{C}\_s(U)$ is undefined (program rejected).

1.5 Computation Graph

For each event (s,t), define $G_s(t)=(V,E)$ where V contains data nodes ($X_s(t)$, μ_s, σ_s, C_s, A_s, D_s, T_s, R_s) and operator nodes ($\circ$, $A\bullet$, $+$, $R\bullet$). Edge $u \rightarrow v$ denotes data-dependence. A computation is *valid* iff every path is confined to variables labeled ind for the same s and never touches forbid.

---

2. Operations

2.1 Baseline Updates (Streaming, Robust)

Given observation $x_{s,i}(t)$ for feature i:

$$\mu_{s,i}(t+1)=(1-\eta_{\mu,i})\,\mu_{s,i}(t)+\eta_{\mu,i}\,x_{s,i}(t),$$

$$\sigma^2_{s,i}(t+1)=(1-\eta_{\sigma,i})\,\sigma^2_{s,i}(t)+\eta_{\sigma,i}\,(x_{s,i}(t)-\mu_{s,i}(t))^2,$$

with η's chosen by Robbins–Monro or bounded adaptives. Median/MAD variants are admissible.

2.2 Adjustment Transform

Let $\circ$ denote Hadamard product. With A_s diagonal and R_s sparse ($|R_{ij}| \le \rho_{max}$):

$$X_s^{adj}(t)=A_s(D_s \circ X_s(t))+T_s+R_s\,X_s(t).$$

Thresholds $T_{s,i}=k_i\,\sigma_{s,i}^{eff}(t)$, with k_i constrained by policy bounds.

2.3 QA / Adversarial Tests

- Physiological bounds: feature-wise intervals.
- Inter-modal coherence: $\sigma^2_{intermodal,s}(t) < \beta \cdot \{\sigma\}^2_{s}(t)$.
- Distributional: univariate KS ($\alpha=0.01$), SPRT for change, Mahalanobis distance using diagonal or robust Σ.
- Low-confidence handling: history weight ≥ 0.7 until convergence.

2.4 Guards and Attestation

- Compile-time: static rejection of $\mathcal{V}\_{forbid}$ or cross-subject references $\Rightarrow$ E-NO-COMPARE.
- Run-time: proxy-risk detector; trigger E-PROXY-RISK if $|corr(input, proxy)|>\tau_p$.
- Load-time: build/schema attestation to ensure protected-class fields absent.

---

3. Proof Obligations and Strategies

3.1 Non-Interference (Bias-Blindness by Construction)

Claim. For any well-typed e over $\mathcal{C}_s(U)$, if worlds W and W′ differ only in demographic attributes or between-group statistics, then ⟦e⟧W = e {W′}. **Strategy.** Induction on typing derivation. Base: variables in $\mathcal{V}_{ind}(s)$ equal across W,W′. Step: each f∈$\mathcal{L}$ is subject-local and references only $\mathcal{V}_{ind}(s)$; forbidden symbols are unreferencable by typing; hence denotation invariant.

3.2 Convergence of Calibration

Assume $F_s(\theta)$ over θ=(μ,σ,C,A,D,T,R) is L-smooth and μ-strongly convex on a compact policy-bounded domain. Projected gradient / coordinate updates with step γ∈(0,2/L) yield [$|\theta^{(t)}_s-\theta^*_s| \le (1-\mu/L)^t |\theta^{(0)}_s-\theta^*_s|$,] *so for error ε:* ($N_{conv} \le (L/\mu)(|\theta^{(0)}_s-\theta^*_s|/\varepsilon)$.)

3.3 EIV Ephemerality

For finite T, $w_{EIV}(t)=\max(0,1-t/T)$ implies $w_{EIV}(t)=0$ for t≥T. Any statistic that depends on E reverts to intra-personal baselines thereafter.

3.4 Information-Utilization Advantage

Let I_{demo} be mutual information from demographic buckets (few bits). Let I_{ind} be the MI from per-subject temporal/multi-modal calibration (tens of bits). With demographic variables excluded by construction, downstream predictions depend only on I_{ind}; ratio $R=I_{ind}/I_{demo}$ typically ≫1 (empirically 6–12×).

3.5 Meta-Learning Safety

Let $\mathcal{D}_{proc}$ contain process-only tuples (timestamp, conv_time, iters, success, eiv_id, reliability). Gradients used to tune hyperparameters are functions of $\mathcal{D}_{proc}$ only. By non-interference, no cross-subject behavioral leakage occurs.

---

4. Complexity

Let $d=d_s$. - Event transform: $A_s(D_s \circ X)$ — O(d); $R_s X$ — $O(nnz(R_s))$ with sparsity s.t. nnz=Θ(d). ⇒ $T_{event}=\Theta(d)$. - QA: KS O(d); Mahalanobis $O(d^2)$ (dense) or O(d) (diagonal/robust); SPRT amortized O(1). - Updates: O(d) time, O(d) space per subject for (μ,σ,C,A,D,T) plus $O(nnz(R_s))$. Fleet space $O(|\mathcal{S}|\cdot d)$.

---

5. Pattern and Paradox Notes (for reviewers)

- **Pattern:** Population-of-one closure behaves as a safety *monoid*: closed under composition; absence of forbidden symbols is absorbing (program invalid), not reflective.

- **Paradox:** Removing a *few* bits (demographics) increases total usable information: architectural removal of brittle, high-variance between-group signals forces exploitation of richer within-subject dynamics ($R \gg 1$).
- **Pattern:** EIV is a scaffolding prior with guaranteed self-destruction; proof is purely algebraic (weights), not trust-based.

---

6. Verification Checklist (Audits)

1) Type audit: SSR satisfied; no $\mathcal{V}$_forbid.
2) Build attestation: schemas contain no protected-class fields.
3) EIV logs: $w\_EIV(t) \to 0$ by $t \geq T$.
4) Proxy checks: $|corr(outputs, any external demographics)| < \tau\_p$ (where lawfully testable).
5) QA thresholds: stability ratios $s\_\{s,i\}(t)$ stabilize; drift triggers local recalibration only.
6) Process isolation: $\mathcal{D}$_proc excludes behavioral measurements.

---

7. Minimal Reference Implementation (pseudo-math)

Per-event: 1) ($x \leftarrow X\_s(t)$); enforce bounds $\rightarrow$ x. 2) $\mu,\sigma \leftarrow Update(\mu,\sigma,x)$. 3) $x\_adj \leftarrow A\_s(D\_s \circ x)+T\_s+R\_s\ x$. 4) Run QA (KS/SPRT/Mahalanobis); if anomaly $\rightarrow$ low-confidence path. 5) Emit *relative-to-baseline* outputs only.

Compile-time: static typecheck ⇒ reject if cross-subject or forbid. **Run-time:** proxy-risk detector; build/schema attested at load.

End (v1.1).