ISSN: 2834-4928

Research Article

Journal of Electrical Electronics Engineering

A Coherence-Gated Bayesian Decision Rule

Raoul Bianchetti[1] and Payam Danesh[2]*

[1]*Information Physics Institute, Gosport, Hampshire, United Kingdom*

[2]*Department of Biosystems Engineering, University of Tehran, Tehran, Iran*

***Corresponding Author**
Payam Danesh, Department of Biosystems Engineering, University of Tehran, Tehran, Iran.

Submitted: 2026, July 17; **Accepted:** 2026, Aug 24; **Published:** 2026, Sep 07

Citation: Bianchetti, R., Danesh, P. (2026). A Coherence-Gated Bayesian Decision Rule. *J Electrical Electron Eng, 5*(5), 01-12.

Abstract
In this paper, we formulate a coherence-gated Bayesian decision rule within Viscous Time Theory (VTT) frame. A standard posterior is multiplied by a bounded gate constructed from alignment, retained coherence cost, informational support, effective informational viscosity, and local sensitivity. The result is normalized only when its total gated mass is positive; otherwise, the rule abstains. We prove boundedness and well-posedness, exact recovery of the Bayesian posterior at the neutral gate, an explicit perturbation bound, an odds decomposition, an exact Bayesian representation through an auxiliary admissibility event, and stability away from zero gated mass. A fully specified three-class synthetic study separates predictive features from an observed support-quality channel. Across convergent, bifurcation, and collapse regimes, coverage decreases from 89.3% to 15.3%, while accuracy on covered cases remains between 98.0% and 99.4%. These values demonstrate the intended selective behavior only under the stated construction.

Keywords: Bayesian Inference, Selective Prediction, Reject Option, Abstention, Viscous Time Theory, Informational Viscosity, Calibration, Uncertainty Quantification

1. Introduction

Bayesian inference separates prior information from the evidential contribution of observed data and combines them through a normalized posterior. The earliest published form is associated with Bayes and Price [1]. Cox later showed how probability can arise from consistency requirements for plausible reasoning [2]. These foundations establish a coherent numerical calculus, but they do not by themselves prescribe whether every computed posterior should be converted into an operational decision.

The distinction between quantified uncertainty and admissible action developed along several lines. Dempster introduced upper and lower probabilities induced by multivalued mappings [3]. Chow derived an optimal recognition rule with a reject action and characterized the error-rejection trade-off [4]. Shafer developed a broader mathematical theory of evidence [5], while Pearl provided a systematic account of probabilistic reasoning in structured intelligent systems [6]. These approaches differ substantially, but each made clear that uncertainty representation and decision policy are related without being identical.

By 2000, information geometry had supplied a differential-geometric language for statistical models, divergences, and local sensitivity [7]. Proper scoring rules then provided principled criteria for evaluating probabilistic forecasts [8]. Work on classification with a reject option replaced ad hoc confidence thresholds with explicit loss-based objectives, and selective classification formalized the trade-off between coverage and conditional risk [9,10]. The central lesson is that withholding a prediction is part of the decision problem and must be evaluated quantitatively, rather than counted as an unqualified gain in accuracy.

Two further developments are directly relevant. Amari's later synthesis clarified the role of geometry in statistical inference and learning [11]. Bissiri, Holmes, and Walker showed that coherent belief updates may be constructed from declared losses, thereby placing non-likelihood reweighting within a generalized Bayesian framework [12]. Modern predictive systems also exposed the practical weakness of treating maximum posterior confidence as a complete reliability measure. Calibration can fail even when classification accuracy is high

[13]. Selective methods for deep networks improved risk-coverage behavior, and integrated reject architectures optimized prediction and selection jointly [14,15]. Large-scale experiments under distribution shift subsequently showed that uncertainty estimates and post-hoc calibration may deteriorate together [16]. Conformal prediction supplies a different route, with finite-sample coverage guarantees under stated exchangeability conditions [17]. These methods therefore provide necessary comparators for any new abstention rule.

Viscous Time Theory (VTT) was proposed as a mathematical description of coherence-dependent temporal response [18]. Related work used coherence weights in structured statistical ensembles while retaining classical counting as a limiting case [19]. This paper adopts only a narrow operational part of that programme. It does not infer a physical law for time from statistical data. Instead, it asks whether VTT-inspired descriptors can define a bounded post-posterior gate with an explicit abstention state.

The research gap is precise. Existing selective prediction methods usually derive selection from confidence, learned risk, loss minimization, conformity scores, or out-of-distribution diagnostics. The VTT proposal uses a declared decomposition into alignment, retained coherence cost, support, viscosity, and local sensitivity. The mathematical status of the normalized gated output, however, must be stated without ambiguity. If the gate is positive, the output can be represented exactly as a conventional Bayesian posterior conditioned on an auxiliary admissibility event. If all gated mass vanishes, normalization is undefined and abstention is mandatory. Thus, the construction does not create a new probability calculus.

The contribution claimed here is limited to four points: first, the VTT diagnostic decomposition is converted into a dimensionless, bounded decision rule. Second, the zero-mass case is treated as an explicit output rather than hidden by arbitrary renormalization. Third, classical consistency, Bayesian representability, common-factor cancellation, and stability are proved. Fourth, a controlled synthetic experiment is specified completely, including its random seed, calibration rule, probability model, degradation mechanism, evaluation metrics, and uncertainty intervals.

2. Mathematical Frame

2.1. Bayesian Baseline

Let $\mathcal{H} = \{H_1, \dots, H_n\}$ be a finite family of mutually exclusive and exhaustive hypotheses. Let E denote observed evidence, let $\pi_i = P(H_i) > 0$ be the prior probability of H_i, and let $L_i(E) = P(E|H_i) \geq 0$ be its likelihood. Assume that $\sum_{j=1}^{n} \pi_j L_j(E) > 0$. The baseline posterior is

$$b_i(E) = P(H_i \mid E) = \frac{\pi_i L_i(E)}{\sum_{j=1}^{n} \pi_j L_j(E)}, \qquad \sum_{i=1}^{n} b_i(E) = 1. \tag{1}$$

Equation (1) is the probability-theoretic baseline. Bayes' theorem does not require conditional independence. Any dependence, latent variable, or temporal structure belongs inside the likelihood model L_i.

Definition 1 (Baseline Support). The posterior support at evidence E is $I_E = \{i : b_i(E) > 0\}$. A gate attached to a hypothesis outside I_E cannot create posterior mass because its baseline mass is zero.

2.2. Structural Descriptors

Let $x = x(E)$ be a dimensionless representation of the evidence in a metric state space (X, d). Each hypothesis H_i is assigned four declared descriptors.

1. The alignment $\Phi_i(x) \in [0,1]$ measures compatibility between the evidence state and the structural region associated with H_i.
2. The retained coherence cost $c_i(x) \in [0,\infty)$ measures the cost assigned to maintaining that evidence-to-hypothesis transition.
3. The informational support $\rho_i(x) \in (0,1]$ measures the quality or completeness of the information supporting the transition.
4. The local sensitivity $\gamma_i(x) \in [0,\infty]$ measures how rapidly the retained cost can change near x.

For a general metric state space, local sensitivity is defined by the upper local slope

$$\gamma_i(x) = \limsup_{y \to x,\, y \neq x} \frac{|c_i(y) - c_i(x)|}{d(y,x)}. \tag{2}$$

If $X \subseteq \mathbb{R}^d$, d is Euclidean distance, and c_i is differentiable at x, Equation (2) reduces to $\gamma_i(x) = \|\nabla c_i(x)\|_2$. For discrete evidence spaces, Equation (2) is replaced by the maximum finite difference over the declared neighborhood graph. This distinction is necessary because a gradient cannot be assumed where no differentiable structure exists.

The effective informational viscosity is defined as the ratio of retained cost to available support:

$$\eta_i(x) = \frac{c_i(x)}{\rho_i(x) + \varepsilon}, \qquad \varepsilon > 0. \tag{3}$$

The regularizer ε prevents division by a small support value from becoming singular. All descriptors are dimensionless. If dimensional descriptors are used in an application, they must be standardized before Equation (3), or the coefficients introduced below must carry the reciprocal units required to make the exponent dimensionless.

Table 1 records the notation used throughout this paper. The table distinguishes probabilistic quantities from diagnostic quantities because confusing them would overstate the mathematical status of the construction.

Symbol	Domain	Definition and role
H_i	hypothesis	The ith mutually exclusive candidate.
E	evidence	Observed data supplied to the probability model and diagnostics.
b_i	[0,1]	Baseline Bayesian posterior from Equation (1).
Φ_i	[0,1]	Coherence-tunnel alignment; an application-specific compatibility score.
c_i	[0,∞)	Retained coherence cost assigned to the transition E → H_i.
ρ_i	(0,1]	Informational support or quality of the transition.
γ_i	[0,∞]	Upper local slope of the retained cost.
η_i	[0,∞)	Effective informational viscosity from Equation (3).
A_i	{0,1}	Hard admissibility indicator.
G_i	[0,1]	Bounded coherence gate.
s_i	[0,1]	Unnormalized gated score.
Λ	[0,1]	Total gated mass; the normalizing constant when positive.
q_i	[0,1]	Normalized gated distribution, defined only for $\Lambda > 0$.

Table 1: Notation and Operational Meaning

2.3. Gate, Score, and Selective Output

Choose thresholds $\phi_{\min}\in[0,1]$, $c_{\max}\in[0,\infty)$, and $\rho_{\min}\in(0,1]$. The hard admissibility indicator is

$$A_i(x) = \mathbf{1}\{\Phi_i(x) \geq \phi_{\min},\ c_i(x) \leq c_{\max},\ \rho_i(x) \geq \rho_{\min}\}. \tag{4}$$

For nonnegative penalty coefficients λ and κ, the bounded gate is

$$G_i(x) = A_i(x)\,\Phi_i(x)\,\rho_i(x)\exp[-\lambda\eta_i(x) - \kappa\gamma_i(x)^2]. \tag{5}$$

The gate multiplies the baseline posterior to form a non-probabilistic score. Its total mass is

$$s_i(E) = b_i(E)G_i\big(x(E)\big), \qquad \Lambda(E) = \sum_{j=1}^{n} s_j\,(E). \tag{6}$$

If $\Lambda\,(E) > 0$, the normalized gated distribution is

$$q_i(E) = \frac{s_i(E)}{\Lambda(E)} = \frac{b_i(E)G_i\big(x(E)\big)}{\sum_{j=1}^{n} b_j\,(E)G_j\big(x(E)\big)}. \tag{7}$$

If $\Lambda(E) = 0$, no normalized output exists. The decision rule therefore returns an abstention state ⊥. For a finite action set with a fixed deterministic tie rule, the selective decision map is

$$\delta(E) = \begin{cases} \arg\max_{1\le i\le n} q_i(E), & \Lambda(E) > 0, \\ \perp, & \Lambda(E) = 0. \end{cases} \tag{8}$$

In VTT terminology, $\Lambda(E) = 0$ is called a collapse-field state. Here that phrase denotes only the operational abstention condition in Equation (8); it does not assert a physical collapse process.

The probabilistic and structural parts of the method is shown in **Figure 1**. The posterior is computed first and is not altered internally. Diagnostics then determine whether the posterior should be reweighted and returned, or whether the procedure should abstain.

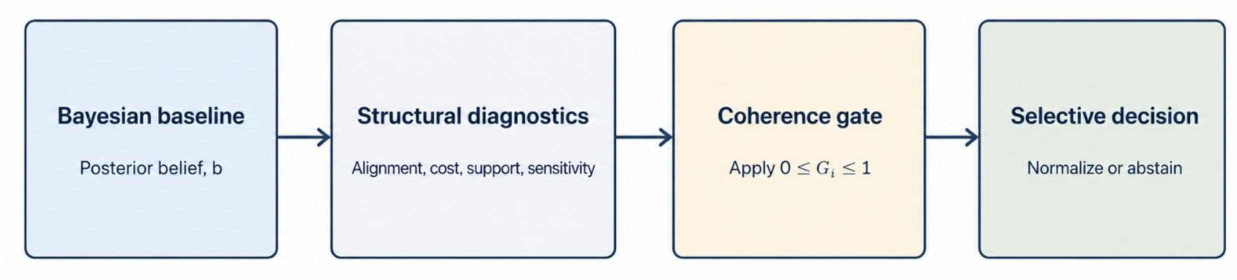

Figure 1: Operational architecture of the coherence-gated Bayesian decision rule. A conventional posterior is computed before any VTT descriptor is applied. The final branch returns either a normalized gated distribution or an explicit abstention state.

2.4. Well-Posedness and Boundedness

Lemma 1 (Gate Bounds). Under the domains specified in Section 2.2 and the conditions $\lambda,\kappa \ge 0$, each gate satisfies $0 \le G_i \le 1$.

Proof. The indicator A_i, alignment Φ_i, and support ρ_i each lie in [0,1]. Equation (3) gives $\eta_i \ge 0$, while Equation (2) gives $\gamma_i^2 \ge 0$. Hence $-\lambda\eta_i - \kappa\gamma_i^2 \le 0$, so the exponential factor belongs to (0,1]. The product of four factors in [0,1] lies in [0,1]. ▫

Theorem 1 (Well-Posed Selective Output). Under the assumptions of Lemma 1,

$$0 \le s_i \le b_i, \qquad 0 \le \Lambda \le 1. \tag{9}$$

If $\Lambda > 0$, the vector $q = (q_1,\dots,q_n)$ in Equation (7) is a probability distribution. Moreover, $\Lambda = 0$ if and only if $G_i = 0$ for every $i \in I_E$.

Proof. By Lemma 1, $0 \le G_i \le 1$. Multiplying by $b_i \ge 0$ yields $0 \le s_i = b_i\, G_i \le b_i$. Summing over i and using $\sum_i b_i = 1$ proves Equation (9). If $\Lambda > 0$, every $q_i = s_i/\Lambda$ is nonnegative and $\sum_i q_i = \Lambda/\Lambda = 1$. If $G_i = 0$ on I_E, then every product $b_i\, G_i$ is zero, including indices outside I_E because $b_i = 0$, so $\Lambda = 0$. Conversely, Λ is a finite sum of nonnegative terms. If the sum is zero, each term $b_i\, G_i$ is zero; for $i \in I_E$, $b_i > 0$, hence $G_i = 0$. ▫

The theorem shows why adding a small constant directly to Λ would be mathematically inappropriate. Such a constant would force a distribution even when every supported transition had failed the declared admissibility test.

2.5. Classical Limit and A Quantitative Error Bound

The neutral gate is the structural state in which every hypothesis carrying posterior mass has full alignment and support, zero retained cost, zero local sensitivity, and passes the hard thresholds. The next theorem proves exact consistency with Bayes and quantifies the deviation from that limit.

Theorem 2 (Classical Consistency). Fix evidence E and its baseline posterior b. Suppose that, for some $0 \le \delta < 1$,

$$\max_{i\in I_E} |G_i - 1| \le \delta.$$

Then $\Lambda \ge 1 - \delta > 0$ and

$$\| q - b \|_1 \le \frac{2\delta}{1-\delta}. \tag{10}$$

Consequently, if a sequence of admissible gates satisfies $G_i^{(m)} \to 1$ for every $i \in I_E$, then $q^{(m)} \to b$ in ℓ_1.

Proof. First,

$$|\Lambda - 1| = \left|\sum_i b_i\,(G_i - 1)\right| \le \sum_i b_i\,|G_i - 1| \le \delta \sum_i b_i = \delta.$$

Therefore $\Lambda \ge 1- \delta > 0$. Next, Equation (7) gives

$$q_i - b_i = \frac{b_i(G_i - \Lambda)}{\Lambda}.$$

Using $|G_i\text{-}\Lambda| \le |G_i\text{-}1|+|1\text{-}\Lambda| \le 2\delta$ and summing yields

$$\| q - b \|_1 = \sum_i \frac{b_i|G_i - \Lambda|}{\Lambda} \le \frac{2\delta}{\Lambda}\sum_i b_i \le \frac{2\delta}{1-\delta}.$$

If $G_i^{(m)} \to 1$ on the finite set I_E, then $\delta_m = \max_{i\in I_E}\left|G_i^{(m)} - 1\right| \to 0$. Equation (10) forces $\|q^{(m)} - b\|_1 \to 0$. ▫

Corollary 1 (VTT Neutral State). If $A_i \to 1$, $\Phi_i \to 1$, $\rho_i \to 1$, $c_i \to 0$, and $\gamma_i \to 0$ for every $i \in I_E$, then $\eta_i \to 0$, $G_i \to 1$, $\Lambda \to 1$, and $q_i \to b_i$.

Proof. Equation (3) gives $\eta_i = c_i/(\rho_i + \varepsilon) \to 0$. Each factor in Equation (5) therefore tends to one. Theorem 2 completes the argument. ▫

2.6. Odds Decomposition and Bayesian Representability

Theorem 3 (Exact Auxiliary-Event Representation). Suppose $\Lambda(E) > 0$. Introduce a binary auxiliary event $Z\in\{0,1\}$ with conditional probabilities

$$P(Z = 1 \mid H_i, E) = G_i(E). \tag{11}$$

Then the gated distribution is the ordinary conditional posterior

$$q_i(E) = P(H_i \mid E, Z = 1). \tag{12}$$

For any i,j with $b_i\, G_i > 0$ and $b_j\, G_j > 0$, its odds satisfy

$$\log\frac{q_i}{q_j} = \log\frac{b_i}{b_j} + \log\frac{G_i}{G_j}. \tag{13}$$

Proof. The law of total probability and Equation (11) give

$$P(Z = 1 \mid E) = \sum_j P\left(Z = 1 \mid H_j, E\right)P\left(H_j \mid E\right) = \sum_j G_j\, b_j = \Lambda.$$

Because $\Lambda > 0$, Bayes' theorem may be applied to the auxiliary event:

$$P(H_i \mid E, Z = 1) = \frac{P(Z = 1 \mid H_i, E)P(H_i \mid E)}{P(Z = 1 \mid E)} = \frac{G_i b_i}{\Lambda} = q_i.$$

Dividing the corresponding expressions for q_i and q_j cancels Λ and gives $q_i/q_j = (b_i/b_j)(G_i/G_j)$. Taking logarithms proves Equation (13). ▫

Theorem 3 Fixes the Scope of the Method. When $\Lambda > 0$, the normalized output is compatible with ordinary probability theory. The gate acts as the conditional probability of an auxiliary admissibility event. When $\Lambda = 0$, the event $Z = 1$ has conditional probability zero, so $P(H_i|E,Z=1)$ is undefined. The abstention rule is therefore not optional bookkeeping; it is the mathematically correct response to conditioning on a zero-probability event in this finite construction.

Corollary 2 (Common-Factor Cancellation). If $G_i(E) = g(E) > 0$ for all $i \in I_E$, then $q_i(E) = b_i(E)$ for every i.

Proof. Equation (7) gives $q_i = b_i\, g/(g\sum_j b_j) = b_i$. ▫

Corollary 2 has an Important Design Consequence. A Diagnostic that Depends only on E and multiplies every hypothesis equally cannot change relative posterior weights. It can still control coverage by setting the common factor to zero, but hypothesis-specific reweighting requires at least one hypothesis-dependent component.

2.7. Stability and Threshold Discontinuity

Theorem 4 (Stability Away from Collapse). Fix a baseline posterior b. Let $g,h \in [0,1]^n$ be two gate vectors satisfying $b \cdot g \geq \alpha$ and $b \cdot h \geq \alpha$ for some $\alpha > 0$. If $Q(g) = b \odot g/(b \cdot g)$, where $\odot$ denotes componentwise multiplication, then

$$\| Q(g) - Q(h) \|_1 \leq \frac{2}{\alpha} \| g - h \|_\infty. \tag{14}$$

Proof. Write $\Lambda_g = b \cdot g$ and $\Lambda_h = b \cdot h$. Add and subtract $b \odot h/\Lambda_g$:

$$\| Q(g) - Q(h) \|_1 \leq \frac{\| b \odot (g-h) \|_1}{\Lambda_g} + \left| \frac{1}{\Lambda_g} - \frac{1}{\Lambda_h} \right| \| b \odot h \|_1.$$

The first numerator is at most $\|g - h\|_\infty \sum_i b_i = \|g - h\|_\infty$. Since $\|b \odot h\|_1 = \Lambda_h$, the second term equals $|\Lambda_h - \Lambda_g|/\Lambda_g$. Also,

$$|\Lambda_h - \Lambda_g| = |b \cdot (h-g)| \leq \| h - g \|_\infty \sum_i b_i = \| h - g \|_\infty.$$

Using $\Lambda_g \geq \alpha$ in both terms proves Equation (14). ▫

The factor $1/\alpha$ shows that normalization becomes ill-conditioned as total gated mass approaches zero. This is not a defect unique to VTT; it is a general property of normalizing a nearly zero nonnegative vector.

Proposition 1 (Hard-Threshold Discontinuity). If A_i is defined by Equation (4), the map from a descriptor to the selective output can be discontinuous at any active threshold.

Proof. Consider one supported hypothesis and vary only its alignment. For $\Phi_i < \phi_{\min}$, $A_i = 0$ and the rule abstains. For $\Phi_i \geq \phi_{\min}$ with all other factors positive, $A_i = 1$ and $\Lambda > 0$. Arbitrarily small changes across $\phi_{\min}$ therefore change the output from $\perp$ to a normalized distribution. ▫

A smooth implementation can replace a hard factor by logistic transitions. For steepness parameters a_ϕ, a_c, $a_\rho > 0$ and logistic function $\sigma(u) = (1+e^{-u})^{-1}$, one option is

$$\tilde{A}_i = \sigma\left(a_\phi(\Phi_i - \phi_{\min})\right)\sigma\left(a_c(c_{\max} - c_i)\right)\sigma\left(a_\rho(\rho_i - \rho_{\min})\right). \tag{15}$$

The smooth gate is differentiable when its descriptors are differentiable, but it removes exact zero-mass abstention unless an additional terminal threshold on Λ is declared. The present paper retains the hard rule so that the abstention state is exact.

Figure 2 shows the hard admissibility region and the resulting jump at the calibrated alignment threshold used later in the synthetic study. The contour field is illustrative: it holds support at one and local sensitivity at zero to isolate the effect of alignment and retained cost.

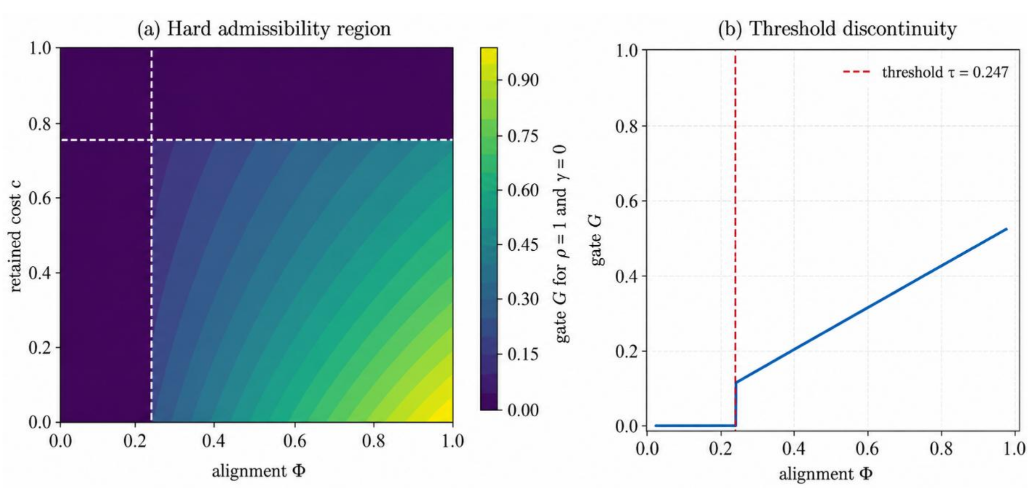

Figure 2: Geometry of the bounded hard gate. Panel (a) shows the admissible region for the illustrative case and ; panel (b) shows the discontinuity introduced by the hard alignment threshold. The figure is a visualization of Equations (4) and (5), not a fitted physical phase diagram.

Example 1 (Calibration is not Automatically Preserved). Suppose a calibrated binary posterior is $b = (0.8,0.2)$ and the gates are $G = (1,0.5)$. Equation (7) gives $q = (0.8,0.1)/0.9=(8/9,1/9)$. Unless the true conditional class probabilities also change after observing the auxiliary event $Z = 1$, q is not calibrated for the original conditioning information E. Thus, classical consistency at $G = (1,1)$ does not imply calibration preservation away from the neutral gate. Calibration must be measured on the covered population.

3. Methods

3.1. Reference Algorithm

Algorithm 1 Implements Equations (1)-(8) without an Implicit Fallback Distribution.

Algorithm 1. Coherence-Gated Bayesian Decision Rule.

1. Compute the baseline posterior $b_i = P(H_i|E)$ for every hypothesis.
2. Compute the declared descriptors Φ_i, c_i, ρ_i, and γ_i using estimators fixed before evaluation.
3. Compute $\eta_i = c_i/(\rho_i + \varepsilon)$ and A_i from Equation (4).
4. Compute G_i from Equation (5) and $s_i = b_i G_i$.
5. Sum the gated mass $\Lambda = \sum_i s_i$.
6. If $\Lambda = 0$, return {status:$\perp$, b, G, Λ, diagnostics}.
7. If $\Lambda > 0$, return {status: covered, $q = s/\Lambda, b, G, \Lambda$, diagnostics}.

The baseline posterior and every diagnostic are retained in the output. This prevents an operator from interpreting q without seeing the probability model that preceded it or the structural factors that changed it.

3.2. Controlled Synthetic Design

The numerical experiment was designed to answer one narrow question: can a support diagnostic that is observed at inference time but omitted from the baseline classifier control abstention in the intended direction? The design does not model a particular physical, medical, or financial system.

There are three classes in $d = 2$ dimensions with equal priors and centers

$$\mu_1 = (-2,0)^\top, \qquad \mu_2 = (2,0)^\top, \qquad \mu_3 = \left(0,2\sqrt{3}\right)^\top. \tag{16}$$

The reference class-conditional model used by the Bayesian classifier is isotropic Gaussian:

$$X \mid H_i \sim \mathcal{N}(\mu_i, \sigma_0^2 I_2), \qquad \sigma_0 = 0.8. \tag{17}$$

Because the priors and covariance determinants are equal, the baseline posterior is computed exactly from standardized squared distances $d_i^2(x) = \| x - \mu_i \|_2^2/\sigma_0^2$:

$$b_i(x) = \frac{\exp[-d_i^2(x)/2]}{\sum_{j=1}^{3} \exp\left[-d_j^2(x)/2\right]}. \tag{18}$$

Each test observation also has a degradation indicator $D \in \{0,1\}$. The class label remains balanced, but the conditional feature variance changes when $D = 1$:

$$X \mid (H_i, D) = \begin{cases} \mathcal{N}(\mu_i, 0.8^2 I_2), & D = 0, \\ \mathcal{N}(\mu_i, 3.5^2 I_2), & D = 1. \end{cases} \tag{19}$$

The three operational regimes differ only in the prespecified degradation probability $p_D = P(D = 1)$: 0.05 for the convergent regime, 0.40 for bifurcation, and 0.85 for collapse. The baseline classifier does not observe D.

An auxiliary support-quality measurement $R \in (0,1)$ is observed by the gate. Conditional on the degradation indicator,

$$R \mid D = 0 \sim \mathrm{Beta}(18,2), \qquad R \mid D = 1 \sim \mathrm{Beta}(2,8). \tag{20}$$

This auxiliary channel is intentionally informative: $E[R|D=0]=0.9$ and $E[R|D=1] = 0.2$. It represents a sensor or preprocessing system that reports data quality separately from the predictive features. The experiment therefore tests the value of using additional structural information; it does not show that the gate can manufacture reliability from the baseline posterior alone.

For the experiment, alignment, retained cost, support, and sensitivity are

$$\Phi_i(x) = \exp\left[-\frac{d_i^2(x)}{2d}\right], \qquad c_i(x) = 1 - \Phi_i(x), \qquad \rho_i(x, R) = R, \tag{21}$$

and, in standardized coordinates $z_i = (x - \mu_i)/\sigma_0$,

$$\gamma_i(x) = \| \nabla_{z_i} c_i \|_2 = \frac{\Phi_i(x)}{d}\sqrt{d_i^2(x)}. \tag{22}$$

The derivative in Equation (22) follows from $c_i = 1 - \exp[-\|z_i\|_2^2/(2d)]$. Differentiation gives $\nabla_{z_i} c_i = \Phi_i z_i / d$; taking the Euclidean norm yields the stated expression.
The experiment uses $\varepsilon = 0.05$, $\lambda = 1$, $\kappa = 0.25$, and $\rho_{\min} = 0.50$. Because $c_i = 1 - \Phi_i$ in this minimal realization, the cost threshold is redundant with the alignment threshold and is not fitted separately. An independent clean calibration sample of 1,000 observations per class fixes

$$\phi_{\min} = Q_{0.05}\left(\max_{1 \le i \le 3} \Phi_i(X_{\text{cal}})\right) = 0.2467273338, \tag{23}$$

where $Q_{0.05}$ denotes the empirical fifth quantile. The threshold is computed before any test regime is generated. Each regime contains 1,000 test observations per class, hence 3,000 observations. All random draws use the NumPy PCG64 generator with seed 20260823. Computations use IEEE 754 double precision.

The experimental choices needed to reproduce the numerical results are listed in **Table 2**. No parameter in the table was optimized on the reported test samples.

Component	**Symbol**	**Value**	**Function in the experiment**
Classes	n	3	Balanced hypotheses with equal prior 1/3.
Dimension	d	2	Euclidean predictive feature space.
Reference deviation	σ_0	0.8	Variance assumed by the baseline classifier.
Degraded deviation	σ_1	3.5	Feature noise when $D = 1$.
Degradation probabilities	p_D	0.05, 0.40, 0.85	Convergent, bifurcation, and collapse regimes.
Calibration size	N_{cal}	3,000	Clean sample used only to fix ϕ_min.
Test size per regime	N	3,000	1,000 independent observations per class.
Alignment threshold	ϕ_{min}	0.2467273338	Fifth empirical quantile from calibration.
Support threshold	ρ_{min}	0.50	Rejects observations with low support measurement.
Viscosity regularizer	ε	0.05	Prevents a small denominator in Equation (3).
Gate penalties	λ, κ	1.00, 0.25	Penalize viscosity and local sensitivity.
Random seed	-	20260823	Fixes calibration and test draws.

Table 2: Prespecified Synthetic Experiment and Gate Parameters

3.3. Evaluation Metrics

Let $a_m = 1\{\Lambda(E_m) > 0\}$ indicate whether test observation m is covered, and let $\hat{y}_m$ be the gated prediction when $a_m = 1$. Empirical coverage is

$$\hat{C} = \frac{1}{N}\sum_{m=1}^{N} a_m. \tag{24}$$

The empirical selective risk under zero-one loss is

$$\hat{R}_{\text{sel}} = \frac{\sum_{m=1}^{N} a_m \, \mathbf{1}\{\hat{y}_m \neq y_m\}}{\sum_{m=1}^{N} a_m}, \qquad \sum_{m=1}^{N} a_m > 0. \tag{25}$$

Selective accuracy is 1- $R\hat{}_{\text{sel}}$. To avoid attributing a gain from abstention to improved classification, the paper also reports unconditional baseline accuracy and the baseline confidence rule at the same empirical coverage. The matched-confidence comparator retains the round($NC\hat{}$) observations with the largest $\max_i b_i$.

For a probability vector $p_m = (p_{m1},\ldots,p_{mn})$, the multiclass Brier score is

$$\mathrm{BS}(p) = \frac{1}{N} \sum_{m=1}^{N} \sum_{i=1}^{n} (p_{mi} - \mathbf{1}\{y_m = i\})^2. \tag{26}$$

Expected calibration error (ECE) uses 15 equal-width confidence bins B_k. With conf (B_k) the mean maximum probability and acc (B_k) the empirical accuracy in bin k,

$$\mathrm{ECE}_{15} = \sum_{k=1}^{15} \frac{|B_k|}{N} |\mathrm{acc}(B_k) - \mathrm{conf}(B_k)|. \tag{27}$$

ECE is reported descriptively because fixed-bin ECE is sample- and bin-dependent. Brier score and ECE for q are evaluated only on covered observations and are compared with the baseline posterior on that same covered subset.

The false activation under low coherence (FALC) diagnostic is the fraction of rejected observations whose maximum baseline posterior is at least 0.7. It measures how often a high-confidence baseline output occurs among cases rejected by the structural rule:

$$\mathrm{FALC}_{0.7} = \frac{\sum_{m=1}^{N} (1 - a_m) \, \mathbf{1}\{\max_i b_{mi} \geq 0.7\}}{\sum_{m=1}^{N} (1 - a_m)}. \tag{28}$$

Risk-coverage curves rank observations by either the structural score $u_m = \min\{R_m, max_i \, \Phi_i(x_m)\}$ or baseline confidence $v_m = \max_i b_i(x_m)$. At each prefix of the descending ranking, selective risk is computed by Equation (25). The area under the risk-coverage curve (AURC) is evaluated by the trapezoidal rule; lower values indicate that errors are concentrated later in the ranking.
For reported binomial proportions, 95% Wilson intervals use z =1.959964. If k successes are observed in N trials and $p\hat{=} k/N$, the interval is

$$\frac{\hat{p} + z^2/(2N)}{1 + z^2/N} \pm \frac{z}{1 + z^2/N} \sqrt{\frac{\hat{p}(1-\hat{p})}{N} + \frac{z^2}{4N^2}}. \tag{29}$$

The intervals quantify sampling uncertainty conditional on the fixed synthetic design. They do not account for model uncertainty, alternative gate definitions, or threshold-selection uncertainty.

3.4. Validation Logic and Failure Tests

The method was checked at four levels. First, numerical assertions were compared with Theorem 1: every computed gate and total mass had to lie in [0,1], and every covered distribution had to sum to one within floating-point tolerance10^{-12}. Second, the neutral gate G_i=1 was verified to return the baseline posterior. Third, the zero-gate vector was verified to return abstention without division. Fourth, the calibrated thresholds were frozen before test generation.

The synthetic study is an application test, not empirical validation of VTT as a physical theory. Its internal validity rests on exact specification and separation of calibration from testing. External validity remains untested.

4. Results

4.1. Analytical Results

The analytical results establish five properties. Theorem 1 proves that the gate and total gated mass remain bounded and that normalization is valid exactly when $\Lambda > 0$. Theorem 2 proves recovery of the Bayesian posterior and supplies the explicit l_1 error bound in Equation (10). Theorem 3 shows that every positive-mass gated output is an ordinary Bayesian posterior conditioned on an auxiliary admissibility event. Corollary 2 proves that a positive common gate cannot change relative posterior weights. Theorem 4 shows Lipschitz stability with respect to the gate whenever total gated mass is bounded below by $\alpha > 0$. Proposition 1 identifies the complementary obstruction: the hard gate is discontinuous at its thresholds.

These statements delimit the framework more strongly than the unqualified claim that the method "extends Bayes." The construction extends the decision pipeline, but its normalized output remains representable within standard conditional probability.

4.2. Synthetic Selective Behavior

The realized degraded fractions were 5.17%, 39.80%, and 84.40% in the convergent, bifurcation, and collapse samples, respectively. As the degraded fraction increased, unconditional baseline accuracy fell from 96.80% to 64.20%. This decline is expected because the classifier continues to use the reference variance$0.8^2\, I_2$ while an increasing share of test features is drawn with variance$3.5^2\, I_2$.

Table 3 reports the decision results. Coverage decreased from 89.27% to 15.27%. Accuracy on covered cases remained above 98%, whereas posterior-confidence selection at the same coverage was lower in every regime. The largest difference occurred in the bifurcation sample: gated selective accuracy was 99.42%, compared with 86.10% for confidence selection at matched coverage. This difference is not evidence that the VTT formula is universally superior. It occurs because the gate observes R, an informative support-quality channel that the confidence rule does not observe.

Regime	Base accuracy, %	Coverage, %	Gated accuracy on covered cases, %	Confidence accuracy at matched coverage, %	FALC0.7, %
Convergent	96.80 (96.11-97.37)	89.27 (88.11-90.32)	99.29 (98.89-99.55)	98.51	88.20
Bifurcation	82.03 (80.62-83.37)	57.57 (55.79-59.32)	99.42 (98.94-99.69)	86.10	93.32
Collapse	64.20 (62.47-65.90)	15.27 (14.02-16.60)	98.03 (96.31-98.96)	79.91	96.34

Table 3: Classification, coverage, and rejection diagnostics(Values in parentheses are 95% Wilson intervals)

The FALC values are high: between 88.20% and 96.34% of rejected observations still have a maximum baseline posterior of at least 0.7. This behavior is a known consequence of extrapolating a closed-set Gaussian classifier. Far from all class centers, relative likelihoods may strongly favor one class even though absolute support is poor. The support channel detects a property that normalized posterior confidence can suppress.

Figure 3 analyzes the full risk-coverage relationship rather than one threshold. Structural ranking has a lower AURC than posterior confidence in the convergent and bifurcation regimes: 0.0027 versus 0.0297, and 0.0507 versus 0.1597, respectively. In the collapse regime, the values are close, 0.2730 versus 0.2746. Once the small high-support subset has been exhausted, neither one-dimensional ranking resolves the ordering among predominantly degraded observations.

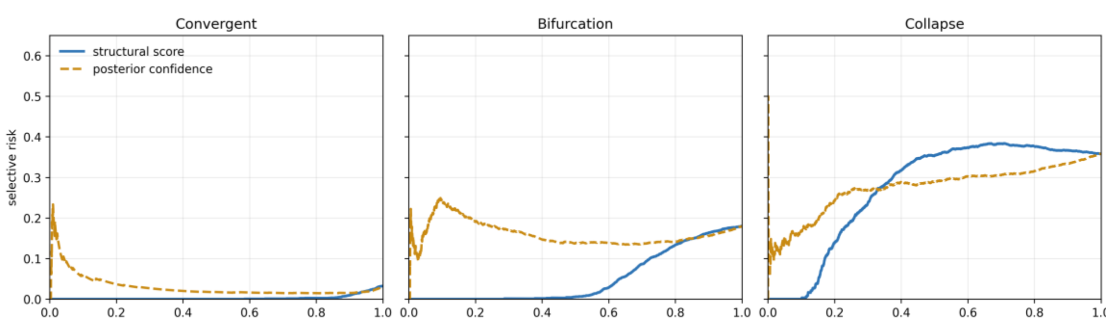

Figure 3: Selective risk as a function of coverage

4.3. Probability Quality on the Covered Population

Table 4 separates selection from probability reweighting. On the covered subset, the baseline posterior is already highly accurate. Multiplying by the hypothesis-dependent gate does not improve Brier score or ECE consistently. In the convergent regime, for example, the covered-case Brier score changes from 0.0122 for b to 0.0142 for q, and ECE changes from 0.0034 to 0.0071. The bifurcation Brier score improves slightly, but its ECE worsens slightly. These mixed results agree with Example 1: gating does not preserve calibration by theorem.

Regime	Base Brier, all	Base ECE, all	Base Brier, covered	Gated Brier, covered	Base ECE, covered	Gated ECE, covered
Convergent	0.0551	0.0187	0.0122	0.0142	0.0034	0.0071
Bifurcation	0.3398	0.1571	0.0118	0.0116	0.0047	0.0058
Collapse	0.6755	0.3291	0.0385	0.0393	0.0222	0.0197

Table 4: Probability-quality metrics before and after gating

The principal numerical effect is therefore selection, not probability correction. A practical implementation should consider returning the baseline posterior together with the gate diagnostics when probability calibration is more important than reweighted class odds.

4.4. Numerical Illustration of the Classical Limit

The classical-limit theorem is independent of the synthetic classification study. To visualize one admissible path, take $b = (0.60,0.30,0.10)$ and positive rates $a = (0.20,0.80,1.50)$. Define

$$G_i(t) = \exp(-a_i t), \qquad q_i(t) = \frac{b_i G_i(t)}{\sum_{j=1}^{3} b_j\, G_j(t)}, \qquad t \geq 0. \tag{30}$$

At $t = 0$, every gate equals one and $q(0) = b$ exactly. For $t > 0$, unequal rates change the odds according to Equation (13). **Figure 4** shows both the ℓ_1 departure from b and the corresponding hypothesis weights. It is a numerical illustration of Theorem 2, not an additional validation dataset.

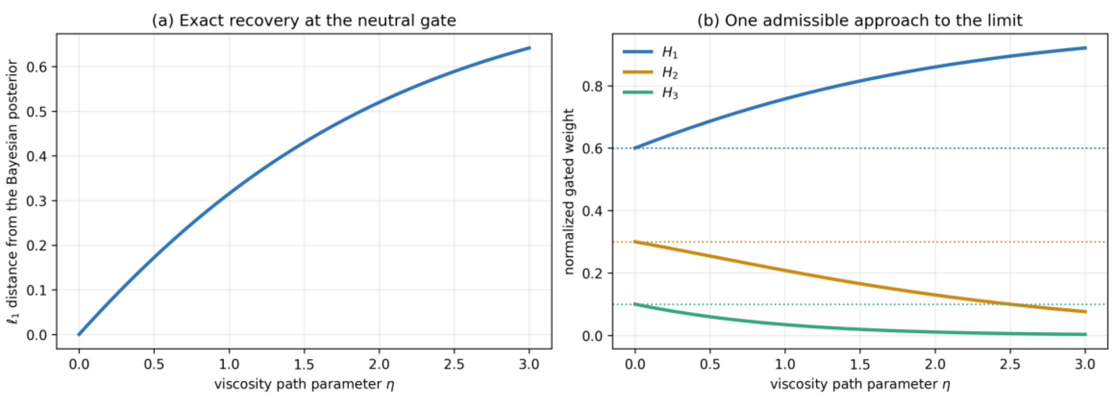

Figure 4: One path through gate space approaching the classical Bayesian limit. Dotted lines are the baseline posterior weights

5. Conclusion

In this paper, we give a mathematically delimited coherence-gated Bayesian decision rule. The gate is bounded, the positive-mass output is a valid distribution, the Bayesian posterior is recovered at the neutral gate, deviations from that limit satisfy an explicit bound, and the normalized result has an exact Bayesian representation through an auxiliary admissibility event. A common positive diagnostic factor cancels from relative posterior weights, while normalization becomes sensitive as total gated mass approaches zero. The hard thresholds create an explicit but discontinuous abstention boundary. The synthetic experiment confirms the intended selective behavior under a fully stated construction. Coverage falls as the observed support channel reports more degraded data, and accuracy on covered cases remains high. This result depends on the auxiliary channel being informative. Probability calibration does not improve consistently after reweighting, so the gate should be evaluated primarily as a selection mechanism unless separate recalibration is performed.

Author Contributions

Conceptualization, R.B. and P.D.; methodology, R.B. and P.D.; formal analysis, P.D.; simulation design, R.B. and P.D.; software and validation, P.D.; visualization, R.B. and P.D.; writing - original draft, R.B. and P.D.; writing - review and editing, R.B. and P.D.; supervision, R.B. All authors approved the manuscript.

Funding

This research received no external funding.

Data and Code Availability

No external empirical dataset was used. The probability model, sample sizes, distributional parameters, calibration rule, numerical seed, metrics, and algorithm required to reproduce the synthetic study are stated in Sections 3.1-3.3 and Table 2. The generated observations contain no personal or proprietary information.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bayes, T., & Price, N. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical transactions of the Royal Society of London*, (53), 370-418.
2. Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American journal of physics, 14*(1), 1-13.
3. Dempster, A. P. (2008). Upper and lower probabilities induced by a multivalued mapping. In Classic works of the Dempster-Shafer theory of belief functions (pp. 57-72). Berlin, Heidelberg: *Springer Berlin Heidelberg*.
4. Chow, C. (1970). On optimum recognition error and reject tradeoff. *IEEE Transactions on information theory, 16*(1), 41-46.
5. Shafer, G. (1976). A Mathematical Theory of Evidence; Princeton University Press: Princeton, NJ, USA. ISBN 978-0-691-10042-5.
6. Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference; Morgan Kaufmann: San Mateo, CA, USA, 1988. ISBN 978-1-55860-479-7.
7. Amari, S. I., & Nagaoka, H. (2000). Methods of information geometry (Vol. 191). American Mathematical Soc.
8. Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association, 102*(477), 359-378.
9. Bartlett, P. L., & Wegkamp, M. H. (2008). Classification with a Reject Option using a Hinge Loss. *Journal of Machine Learning Research, 9*(8).
10. El-Yaniv, R. (2010). On the Foundations of Noise-free Selective Classification. *Journal of Machine Learning Research, 11*(5).
11. Amari, S. I. (2016). Information geometry and its applications. *Springer.*
12. Bissiri, P. G., Holmes, C. C., & Walker, S. G. (2016). A general framework for updating belief distributions. Journal of the Royal Statistical Society Series B: *Statistical Methodology, 78*(5), 1103-1130.
13. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017, July). On calibration of modern neural networks. In International conference on machine learning (pp. 1321-1330). PMLR.
14. Geifman, Y., & El-Yaniv, R. (2017). Selective classification for deep neural networks. *Advances in neural information processing systems, 30.*
15. Geifman, Y., & El-Yaniv, R. (2019, May). Selectivenet: A deep neural network with an integrated reject option. In International conference on machine learning (pp. 2151-2159). *PMLR.*
16. Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., ... & Snoek, J. (2019). Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. Advances in neural information processing systems, 32.
17. Angelopoulos, A. N., & Bates, S. (2023). Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning, 16*(4), 494-591.
18. Danesh, P., & Bianchetti, R. (2026). Viscous Time Theory (VTT): A mathematical proposal for coherence-dependent temporal response. *Journal of Applied Mathematics and Physics, 14*(6), 2307-2316.
19. Danesh, P., & Bianchetti, R. (2026). Coherence-Weighted Entropy for Structured Statistical Ensembles: A Conservative Extension of Boltzmann Counting. *Available at SSRN 6783683*